

**Dinâmica do Mercado Imobiliário: Fatores Determinantes no Valor das Habitações em Portugal**

Maria João Alves Ferreira

06/2021

Maria João Alves Ferreira. **Dinâmica do Mercado Imobiliário: Fatores Determinantes no Valor das Habitações em Portugal**

**Dinâmica do Mercado Imobiliário: Fatores Determinantes no Valor das Habitações em Portugal**

Maria João Alves Ferreira

06/2021

Este trabalho não inclui as críticas e sugestões feitas pelo Júri



**MESTRADO**

MÉTODOS DE APOIO À DECISÃO EMPRESARIAL

# **Dinâmica do Mercado Imobiliário: Fatores Determinantes no Valor das Habitações em Portugal**

Maria João Alves Ferreira

Orientador(es):

Aldina Isabel de Azevedo Correia

Vanda Marlene Monteiro Lima

# Dedicatória

Aos meus Pais, estejam onde estiverem, estarão certamente orgulhosos do meu percurso.  
Ao meu marido e filho.



# Resumo

A habitação é o principal ativo da maioria das famílias, pelo que a volatilidade dos preços pode afetar a decisão de compra dos potenciais proprietários.

O presente estudo pretende deste modo auxiliar na tomada de decisão ao nível individual e ao nível dos decisores políticos nos aspetos relacionados com a habitação em Portugal. Ao nível individual, pretende servir de meio de auxílio na decisão de compra/investimento, quer do comprador, como do investidor. Ao nível dos decisores políticos, quer sejam de âmbito municipal, intermunicipal ou nacional, como ajuda na tomada de decisão sobre as melhores políticas de promoção, renovação e reabilitação da habitação nas suas áreas de intervenção.

Para este trabalho foram utilizados dados disponíveis no PORDATA – Base de Dados Portugal Contemporâneo, de todos os 308 municípios de Portugal. A metodologia adotada consiste na utilização de uma regressão linear múltipla, no sentido de perceber quais os fatores que mais influenciam o valor médio das habitações transacionadas por município. Posteriormente efetuou-se uma análise no sentido de identificar possíveis *clusters* dentro dos municípios, tendo em conta estes fatores.

Os factores considerados incluem variáveis relativas à Localização dos municípios, Densidade Populacional dos municípios, Poder de Compra dos habitantes dos municípios, Dimensão média das Famílias dos municípios, Desemprego nos municípios, Qualidade Ambiental nos municípios, Envelhecimento nos municípios, Nível médio de Educação nos municípios, Estado Civil dos habitantes dos municípios, Habitação Própria pelo proprietário nos municípios, Habitações existentes nos municípios. Os resultados sugerem que apenas a densidade populacional e a qualidade ambiental não têm influência significativa no valor médio por município das habitações transacionadas. Podem ainda considerar-se 2, 3 ou 4 *clusters* de municípios, destacando-se sempre um grupo de municípios com o VMPT muito mais alto do que os outros.

**Palavras-chave:** Valor das Habitações, Portugal, Regressão Linear Múltipla, Análise de *Clusters*.



# Abstract

Housing is the main asset of most families, therefore, price volatility can affect potential owners' purchasing decision.

The present study aims to support decision-making at individual level and at the level of political decision-makers at the level of housing in Portugal. At the individual level, it intends aid in the purchase/investment decision, either of the buyer, as well as the investor. At the level of political decision makers, whether they are municipal, inter-municipal or national, it intends to be an aid in decision-making on the best policies for the promotion, renewal and rehabilitation of housing in their intervention areas.

In this work data available in PORDATA – Portugal Contemporary Database of all 308 municipalities in Portugal were used. The methodology adopted consists in a multiple linear regression, in order to understand which factors most influence the average value of the dwellings traded by municipality. Subsequently, an analysis was carried out in order to identify possible clusters within the municipalities.

Factors considered include variables related to Location of municipalities, Population Density of the municipalities, Purchasing power of the inhabitants of the municipalities, Average size of families in the municipalities, Unemployment in municipalities, Environmental Quality in municipalities, Aging in municipalities, Average level of education in municipalities, Civil status of the inhabitants of the municipalities, Own housing by the owner in the municipalities, Existing housing in municipalities. The results suggest that only population density and environmental quality have no significant influence on the average value per municipality of transacted housing.

2, 3 or 4 clusters of municipalities can be considered, always highlighting a group of municipalities with a much higher housing price than the others.

**Keywords:** Housing Price, Portugal, Multiple Linear regression, Cluster Analysis.



# Agradecimentos

Este projeto não teria sido possível sem a colaboração de diversas pessoas, que tiveram uma intervenção direta na elaboração do mesmo.

Não podia deixar de expressar os meus agradecimentos ao meu marido e filho, onde fui buscar forças para continuar.

À minha amiga Raquel Francisco, que me apoiou nos momentos de stress, cansaço e desânimo e me incentivou para terminar.

Às colegas de curso Raquel, Mónica, Marta, Rosa e Fátima, obrigada por tudo, foi um percurso bonito.

Gostaria de agradecer às minhas orientadoras Professora Aldina Correia e Professora Vanda Lima que me apoiaram, me deram sugestões de como orientar o projeto e me chamaram à realidade quanto aos objetivos a cumprir, nunca desistindo.

Por fim os meus agradecimentos à ESTG-Escola Superior de Tecnologia e Gestão, instituição onde completei a licenciatura, e às coordenadoras do mestrado, nomeadamente a Professora Aldina Correia que me desafiou a entrar nesta aventura.

Obrigada aos restantes professores que me facultaram as ferramentas e materiais necessários, que me permitiram adquirir os conhecimentos necessários para concluir mais uma etapa.

Muito obrigada a todos.



# Conteúdo

<b>Lista de Figuras</b>	<b>xi</b>
<b>Lista de Tabelas</b>	<b>xiii</b>
<b>Lista de Siglas</b>	<b>xv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Apresentação e Oportunidade do Tema . . . . .	1
1.2 Objetivos . . . . .	2
1.3 Contributos inovadores . . . . .	2
1.4 Estrutura . . . . .	3
<b>2 Revisão de Literatura</b>	<b>5</b>
2.1 Mercado Imobiliário . . . . .	5
2.2 Análise Bibliométrica . . . . .	7
2.3 Fatores Determinantes para o Valor das Habitações Transacionadas . . . . .	9
2.3.1 Localização . . . . .	9
2.3.2 Densidade Populacional . . . . .	10
2.3.3 Poder de compra . . . . .	11
2.3.4 Dimensão das famílias . . . . .	12
2.3.5 Desemprego . . . . .	12
2.3.6 Qualidade Ambiental . . . . .	13
2.3.7 Envelhecimento . . . . .	13
2.3.8 Nível de Educação . . . . .	14
2.3.9 Estado Civil . . . . .	15
2.3.10 Habitação Própria . . . . .	15
2.3.11 Habitações existentes . . . . .	16
2.3.12 Síntese da Revisão de Literatura . . . . .	17
2.4 Métodos Quantitativos de Apoio à Decisão . . . . .	17
2.4.1 Regressão linear . . . . .	17
Multicolinearidade . . . . .	19
Análise de Resíduos . . . . .	19
2.4.2 Análise de Clusters . . . . .	20
<b>3 Metodologia e Resultados</b>	<b>23</b>
3.1 Metodologia . . . . .	23
3.2 Evolução do VMPT em Portugal . . . . .	23
3.3 Dados . . . . .	25
3.3.1 Variável resposta . . . . .	26
3.3.2 <i>Proxys</i> para os fatores que influenciam o valor das habitações . . . . .	26
3.4 Fatores que influenciam o VMPT em Portugal . . . . .	28

3.4.1	Modelo 1 . . . . .	28
3.4.2	Validação de Pressupostos e Análise de Resíduos . . . . .	30
3.4.3	Modelo 2 . . . . .	32
3.4.4	Validação de Pressupostos e Análise de Resíduos . . . . .	34
3.5	Identificação de <i>Clusters</i> de Municípios Homogéneos . . . . .	37
<b>4</b>	<b>Conclusões, Contribuições, Limitações e Trabalho Futuro</b>	<b>45</b>
	<b>Bibliografia</b>	<b>49</b>
<b>A</b>	<b>Distribuição dos Municípios por NUTS III</b>	<b>53</b>
<b>B</b>	<b>Dendrograma</b>	<b>55</b>
<b>C</b>	<b>Comparação entre médias por Cluster do VMPT e das outras variáveis</b>	<b>57</b>
<b>D</b>	<b>Testes ANOVA para 3 e 4 Clusters</b>	<b>61</b>
<b>E</b>	<b>Elementos do Cluster 2, para <math>k = 2</math></b>	<b>63</b>
<b>F</b>	<b>Elementos do Cluster 2 e 3, para <math>k = 3</math></b>	<b>65</b>
<b>G</b>	<b>Elementos do Cluster 1, 3 e 4, para <math>k = 4</math></b>	<b>69</b>

# Lista de Figuras

2.1	Publicações Tema Housing Prices Fonte: <i>Web of Science</i> . . . . .	8
2.2	Autores com Maior Número de Publicações Fonte: <i>Web of Science</i> . . . . .	8
2.3	Países com Maior N. <sup>o</sup> de Publicações Fonte: <i>Web of Science</i> . . . . .	9
3.1	Evolução do VMPT em Portugal . . . . .	24
3.2	Evolução do VMPT por NUTSII . . . . .	25
3.3	Histograma e <i>P – PPlot</i> . . . . .	30
3.4	<i>Scatterplot</i> . . . . .	31
3.5	Gráfico dos <i>Outliers</i> . . . . .	32
3.6	Histograma e <i>P-P Plot</i> . . . . .	35
3.7	Convergência . . . . .	38
3.8	Testes ANOVA para $k = 2$ clusters . . . . .	38
3.9	Representação das medidas de validação . . . . .	41
3.10	Representação dos centróides finais para $k = 2$ . . . . .	42
3.11	Número de Municípios por <i>cluster</i> . . . . .	42
3.12	Médias dos VMPT por <i>cluster</i> . . . . .	43
B.1	Dendrograma . . . . .	56
C.1	Médias por <i>cluster</i> para $k = 3$ . . . . .	58
C.2	Médias por <i>cluster</i> para $k = 4$ . . . . .	59
D.1	Testes ANOVA para $k = 3$ clusters . . . . .	61
D.2	Testes ANOVA para $k = 4$ clusters . . . . .	61



# Lista de Tabelas

2.1	Fatores, sinal esperado e autores . . . . .	16
2.2	Etapas Análise Clusters Fonte: Reis (2001) . . . . .	21
3.1	Distribuição dos Municípios por NUTS I . . . . .	25
3.2	Distribuição dos Municípios por NUTS II . . . . .	26
3.3	Descrição das Variáveis Consideradas . . . . .	26
3.4	Cálculo de Variáveis . . . . .	27
3.5	Estatística Descritiva das Variáveis . . . . .	27
3.6	Teste ANOVA . . . . .	28
3.7	Coeficientes do modelo de regressão . . . . .	29
3.8	Diagnóstico de Colinearidade . . . . .	31
3.9	Matriz de Correlações . . . . .	33
3.10	Modelo de Estimação OLS . . . . .	34
3.11	Modelo de Estimação GLS . . . . .	36
3.12	Confirmação das Hipóteses . . . . .	37
3.13	Medidas de Validação dos <i>Clusters</i> . . . . .	39
3.14	Optimal Scores . . . . .	40
3.15	Centróides Finais dos Clusters . . . . .	40
A.1	Distribuição dos Municípios por NUTS III . . . . .	54
E.1	Municípios no <i>Cluster 2</i> , para $k = 2$ . . . . .	63
F.1	Municípios no <i>Clusters 2 e 3</i> , para $k = 3$ . . . . .	66
F.2	Municípios no <i>Clusters 2 e 3</i> , para $k = 3$ . . . . .	67
G.1	Municípios no <i>Cluster 1, 3 e 4</i> , para $k = 4$ . . . . .	69
G.2	Municípios no <i>Cluster 1, 3 e 4</i> , para $k = 4$ . . . . .	70



# Lista de Siglas

<b>GLM</b>	<b>G</b> eneralized <b>L</b> inear <b>M</b> odel
<b>INE</b>	Instituto <b>N</b> acional de <b>E</b> statística
<b>IPH</b>	Índice de <b>P</b> reços da <b>H</b> abituação
<b>EUA</b>	<b>E</b> stados <b>U</b> nidos da <b>A</b> mérica
<b>OCDE</b>	<b>O</b> rganização para a <b>C</b> ooperação e <b>D</b> esenvolvimento <b>E</b> conómico
<b>PORDATA</b>	Base de <b>D</b> ados de <b>P</b> ortugal Contemporâneo
<b>STAR</b>	<b>S</b> ituation, <b>T</b> ask, <b>A</b> ction, <b>R</b> esults
<b>VMPT</b>	<b>V</b> alor <b>M</b> édio dos <b>P</b> rédios <b>T</b> ransacionados
<b>PIB</b>	<b>P</b> roduto <b>I</b> nterno <b>B</b> ruto
<b>SPSS</b>	<b>S</b> tatistical <b>P</b> ackage for the <b>S</b> ocial <b>S</b> ciences
<b>APN</b>	<b>A</b> verage <b>P</b> roportion of <b>N</b> on-overlap
<b>AD</b>	<b>A</b> verage <b>D</b> istance
<b>ADM</b>	<b>A</b> verage <b>D</b> istance between <b>M</b> eans
<b>FOM</b>	<b>F</b> igure <b>O</b> f <b>M</b> erit
<b>BHI</b>	<b>B</b> iological <b>H</b> omogeneity <b>I</b> ndex
<b>BSI</b>	<b>B</b> iological <b>S</b> tability <b>I</b> ndex
<b>OLS</b>	<b>O</b> rdinal <b>L</b> east <b>S</b> quares
<b>GLS</b>	<b>G</b> eneralized <b>L</b> east <b>S</b> quares
<b>IBM</b>	<b>I</b> nternational <b>B</b> usiness <b>M</b> achines Corporation
<b>NUTS</b>	<b>N</b> omenclatura das <b>U</b> nidades <b>T</b> erritoriais para fins estatísticos
<b>BCE</b>	<b>B</b> anco <b>C</b> entral <b>E</b> uropeu
<b>FMI</b>	<b>F</b> undo <b>M</b> onetário <b>I</b> nternacional
<b>TLC</b>	<b>T</b> eorema do <b>L</b> imite <b>C</b> entral
<b>VIF</b>	<b>V</b> ariance <b>I</b> nflation <b>F</b> actor
<b>GIS</b>	<b>G</b> eografic <b>I</b> nformation <b>S</b> ystem



# Capítulo 1

## Introdução

### 1.1 Apresentação e Oportunidade do Tema

As políticas de habitação e as dinâmicas habitacionais são um tema de grande complexidade sobretudo atualmente devido aos inconstantes preços das habitações, resultado da incerteza das economias nacionais e internacionais.

A crise financeira iniciada em 2007 foi desencadeada por proprietários com excesso de alavancagem e uma forte crise no mercado imobiliário dos EUA (Duffie 2019). Na sua linha de raciocínio, este autor (Duffie 2019), acrescenta que um sistema financeiro bem supervisionado teria sido muito mais resistente a este e a outros tipos de choques severos. Ao contrário disso os intermediários financeiros faliram, ou foram resgatados, ou reduziram de forma drástica o seu fornecimento de liquidez e crédito à economia. Resumidamente, o sistema financeiro central deixou de desempenhar as funções pretendidas para a economia real com um nível razoável de eficácia. Como resultado, o impacto do choque do mercado imobiliário sobre o resto da economia foi muito maior do que o esperado. Refere ainda alguns avanços positivos efetuados no pós-crise, nomeadamente, na melhoria da capitalização das maiores instituições financeiras, uma redução de práticas inseguras e infraestruturas nos mercados de financiamento de títulos e derivados, e uma presunção significativamente reduzida de que as maiores empresas financeiras serão resgatadas pelo dinheiro do contribuinte no futuro.

Com as profundas alterações sociais do final do século XX, surgiram novas necessidades no setor imobiliário. Questões como as transformações da estrutura demográfica, relação casa/trabalho, alterações nas dinâmicas socioculturais estão intimamente relacionadas com o aumento da procura. O ser humano (consumidor) vive na busca da satisfação dos seus desejos e objetivos. Com as alterações sentidas na sociedade, o consumidor passou a ser mais exigente, intolerante e com vontade de adquirir valor agregado nos produtos (Kotler e Armstrong 2000). Daí a dificuldade cada vez maior de prever o comportamento dos consumidores, uma vez que cada um terá uma forma diferente de interagir perante os estímulos exteriores. Torna-se imperioso o estudo destes comportamentos para conhecimento das suas preferências, expectativas e desejos, no sentido de poder oferecer produtos e serviços que os satisfaçam. É certo que o consumidor tem gostos e desejos ilimitados, no entanto apenas conseguirá satisfazer alguns, pelo que opta por aqueles que maximizarão o seu prazer. Há uma estreita ligação entre qualidade de produtos e serviços, satisfação de consumidores e lucros da empresa. Níveis mais elevados de qualidade resultam em níveis mais elevados de satisfação de clientes, ao mesmo tempo em que sustentam preços mais altos e (frequentemente) custos menores (Kotler e Armstrong 2000).

De acordo com Abraham Maslow, criador da mais conhecida teoria sobre as necessidades, citado por Hoffman (1988), a compra de habitação está englobada nas necessidades de segurança, que se encontram na segunda linha de prioridades do consumidor, logo após a satisfação das necessidades fisiológicas.

A habitação é um bem de consumo essencial, diferenciado e duradouro, é um indicador indireto de *status* e de distinção dos rendimentos dos consumidores. Trata-se, portanto, do principal ativo para a maioria das famílias, onde são gastos a maior parte dos seus recursos, pelo que, quando se decidem pela compra, as pessoas não têm apenas em consideração as características intrínsecas do ativo, mas também fatores como vizinhança, acessibilidade a serviços ou qualidade ambiental (Calmasur 2016).

Duffie (2019) aponta para alguns avanços positivos significativos que foram feitos após a crise de Setembro de 2008, agudizada pela falência Lehman Brothers, que podem ter um impacto positivo na compra/venda de habitações. O desequilíbrio entre procura e oferta habitacional é um dos determinantes-chave dos movimentos dos preços da habitação (Chong e X. Li 2019), criando o mercado imobiliário.

## 1.2 Objetivos

O objetivo principal deste trabalho é (O1) determinar os fatores que influenciam o valor das habitações nos municípios em Portugal. Trata-se de um estudo econométrico com o objectivo de avaliar qual o impacto de diversas características dos municípios no valor médio das transações de habitação, praticado nestes municípios. Pretende-se, ainda, compreender de que forma a variação destas características municipais influencia a evolução desse valor. Além disso, ainda são objetivos deste trabalho (O2) estudar o comportamento das oscilações de valor da habitação em Portugal e (O3) determinar as semelhanças entre municípios no que respeita a estas oscilações.

## 1.3 Contributos inovadores

Alguns estudos foram já elaborados sobre os determinantes que afetam o valor da habitação.

No entanto, o foco desses estudos centrou-se no âmbito internacional, outros mais regional ou acerca da proximidade a grandes centros, mas a literatura é escassa ao nível de municípios, sendo desta forma este trabalho um contributo inovador também em termos teóricos.

O que se pretende neste trabalho é fazer uma análise a nível nacional com carácter intra-municipal, tentando distinguir os municípios portugueses de acordo com o Valor médio dos prédios transacionadas (VMPT), identificando quais os fatores que o influenciam, incluindo fatores sociais, características da população e características ambientais, entre outros.

Este estudo fornece, então, contributos teóricos e práticos para a tomada de decisão, quer ao nível individual, por parte dos compradores/investidores do mercado imobiliário residencial, quer ao nível colectivo, por parte dos decisores políticos na definição das políticas públicas de habitação.

## **1.4 Estrutura**

A presente dissertação está organizado em cinco Capítulos. No Capítulo 1 faz-se uma apresentação do tema a tratar, dos objetivos e principais contributos e descreve-se a presente estrutura. No Capítulo 2 é feita uma revisão de literatura, focando nomeadamente os determinantes do valor das habitações identificados pelos investigadores. De seguida, no Capítulo 3, apresenta-se a metodologia e os resultados empíricos. O trabalho termina com o Capítulo 4 que inclui as principais conclusões, as limitações da abordagem apresentada e algumas propostas de trabalhos futuros.



## Capítulo 2

# Revisão de Literatura

### 2.1 Mercado Imobiliário

É inegável a importância da posse ou usufruto de uma habitação. Esta necessidade é de forma imperceptível regulada pelo mercado imobiliário, que medeia as transações entre os que procuram casa e os que a oferecem para venda. O mercado imobiliário não é uma entidade uniforme, mas pode ser considerado como um conjunto distinto de sub-mercados, baseados na estrutura e atributos da localização dos bairros e residências (Adair, Berry e W. S. McGreal 1996). As marcas deixadas pela crise de 2008 (que ganhou evidência a partir do verão de 2007), são evidentes no crédito à habitação, tanto do lado das instituições financeiras, como dos próprios clientes, de acordo com Carlos Vintém, diretor comercial da União de Créditos Imobiliários de Portugal (Idealista 2019). O controlo passou a ser mais apertado pelo setor bancário, e passou a existir mais resistência no acesso ao crédito por parte dos consumidores, dificultando assim a compra e venda de imóveis, diminuindo a atividade imobiliária. Segundo dados do INE<sup>1</sup>, o mínimo do IPH - Índice de preços da habitação<sup>2</sup>, índice que mede a evolução dos preços dos alojamentos/fogos no mercado residencial do território nacional, atingiu o seu valor mínimo no 2.º Trimestre de 2013, tendo vindo a subir a partir daí.

De acordo com W. Peng (2003), o mercado imobiliário desempenha um papel muito importante na economia de Hong Kong. A habitação é a mais importante forma de poupança para muitas das famílias. O autor realça o fato de metade do crédito interno estar associado a empréstimos hipotecários para a compra de propriedades residenciais privadas, para a construção e para a promoção imobiliária. As mudanças nos preços e rendas dos imóveis influenciam o consumidor, inflacionando os preços que afetam a competitividade de Hong Kong, baseada numa economia de serviços. A acrescentar a tudo isto a venda de terrenos e o imposto de selo sobre transações imobiliárias também têm sido uma fonte significativa de receita do governo.

Tsatsaronis e Zhu (2004) afirmam que a habitação apresenta um conjunto de características distintas face a outros ativos, que podem explicar a ligação entre o desempenho económico e o mercado imobiliário residencial. Fatores como o aumento do salário disponível das famílias, mudanças graduais nos dados demográficos (como o tamanho relativo das gerações mais velhas e mais novas), características permanentes do sistema tributário que podem encorajar a compra de casa própria em oposição a outras formas de acumulação de riqueza,

---

<sup>1</sup>INE – Instituto Nacional de Estatística, disponível em [https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_destaquas&DESTAQUESdest\\_boui=281088048&DESTAQUESmodo=2](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaquas&DESTAQUESdest_boui=281088048&DESTAQUESmodo=2)

<sup>2</sup>IPH – baseia-se na estimação de uma relação funcional entre o logaritmo dos preços de transação dos alojamentos e as suas características (e.g., área, localização),

o nível médio de taxas de juros (possivelmente relacionadas ao comportamento de longo prazo da inflação), o custo do terreno, o custo de construção e investimentos na melhoria da qualidade do parque habitacional existente são determinantes da oferta de habitação (Tsatsaronis e Zhu 2004).

Gupta e Kabundi (2010) no seu estudo de 2010, sobre o impacto da política monetária sobre a inflação dos preços imobiliários em nove divisões da economia dos Estados Unidos da América (EUA), baseado em séries temporais, concluíram que, em geral, a inflação dos preços da habitação responde negativamente a um choque positivo de política monetária.

Song e Knaap (2003) através de um estudo efetuado em vários bairros de Washington County, Oregon, desenvolveram várias medidas da forma urbana, e estas foram usadas para caracterizar os diversos bairros. As medidas evidenciam diferenças significativas no caráter do design dos bairros urbanos. Mais importante, os resultados mostraram que o valor da casa depende das características particulares de design que ela tem a oferecer. Resumidamente, o design é importante para definir o preço das habitações. Além disso, embora o preço de uma casa unifamiliar seja difícil de prever, as relações estatísticas entre as características de design nas 48 000 observações de vendas são robustas. Com base em sinais de mercado, os resultados sugeriram que as virtudes dos novos bairros urbanistas incluem conectividade interna aprimorada, bons caminhos pedestres, acesso a lojas comerciais, bons acessos de trânsito e menos conectividade externa.

A habitação constitui um direito fundamental do Homem, tipificado no Artigo 65.<sup>o</sup> (Habitação e urbanismo) da Constituição da República Portuguesa (República Portuguesa 1976). Trata-se, portanto, de um bem essencial à subsistência humana e o principal bem da maioria das famílias.

Com base nos dados divulgados pelo INE, Portugal tinha em 2017 cerca de 3,6 milhões de edifícios de habitação familiar clássica e 5,9 milhões de alojamentos familiares clássicos, o que representa um acréscimo de 0,19% e 0,16% respetivamente, face ao ano anterior. O número de edifícios licenciados em Portugal cresceu 9,9% em 2017 (+11,4% em 2016), correspondendo a 18 621 edifícios. Em 2017 registou-se um acréscimo de 20,6% no número de alojamentos familiares transacionados em relação ao ano anterior, tendo assim ultrapassado o registo máximo de vendas de habitações observado em 2010 (129 950). Segundo o Relatório de Estabilidade Financeira do Banco de Portugal (Banco de Portugal 2019), a dinâmica de preços do mercado imobiliário residencial em Portugal tem sido determinada pela melhoria do rendimento das famílias, pelo baixo nível das taxas de juro e pelo menor grau de restritividade dos critérios de concessão de crédito à habitação. A procura por não residentes, em parte associada a autorizações de residência, tem também dinamizado alguns segmentos deste mercado. Por sua vez, a forte dinâmica do setor do turismo, em particular do alojamento local, tem contribuído para a evolução dos preços do imobiliário residencial (Banco de Portugal 2019).

No seu relatório de Junho de 2020 o Banco de Portugal (2020) prevê que a crise pandémica relacionada com a COVID-19, em particular nas atividades ligadas ao turismo, poderá afetar negativamente a evolução dos preços do mercado imobiliário. Durante o ano de 2019, este segmento continuou a apresentar dinamismo, embora se tenha observado alguma moderação na evolução dos preços, com uma taxa de variação média de 8,4%, 0,5 pontos percentuais inferior à do ano anterior. As transações registaram uma desaceleração, em termos homólogos, em 2019, tanto em número como em montante de vendas de alojamentos familiares, de forma transversal às várias regiões, com destaque para a queda do número de transações nas

regiões Norte (incluindo a área metropolitana do Porto) e Algarve e a variação virtualmente nula na área metropolitana de Lisboa.

A acrescentar a isto, verificou-se o aumento do número de fogos licenciados, o que se poderá traduzir no aumento da oferta de alojamentos novos para habitação, colocando assim pressão descendente sobre os preços do imobiliário residencial (Banco de Portugal 2020).

Neste contexto, o Banco de Portugal (2020) prevê que a redução da atividade económica, o impacto sobre o rendimento dos particulares, a deteriorização da confiança dos consumidores, e a incerteza em relação à crise pandémica e económica poderá contribuir para uma correção dos preços neste setor. Desta forma, é convicção do Banco de Portugal (2020) que a lenta recuperação económica a nível mundial deverá traduzir-se em menores fluxos de turismo externo, fator que nos últimos anos muito contribuiu para o aumento dos preços no mercado imobiliário residencial. As medidas tomadas pelo governo nomeadamente no regime de moratórias, públicas e privadas, no crédito à habitação, tentando prevenir o incumprimento e subsequente execução do colateral, deverão contribuir no curto-prazo, para diminuir os impactos sobre este mercado. A manutenção do ambiente de baixas taxas de juro e as taxas de rendibilidade elevadas face a outros ativos poderão funcionar como um atenuante sobre estas pressões. Contudo poderão existir impactos negativos mais persistentes associados à crise sanitária, como uma menor procura por parte de não residentes, bem como por edifícios de escritórios devido à generalização do trabalho remoto nos serviços (Banco de Portugal 2020).

A diversidade de atributos da habitação: duração indefinidamente longa, características físicas variadas, dimensões várias e impossibilidade de deslocação geográfica, originam juízos diferentes dos agentes económicos, pelo que cada imóvel terá um valor diferente para cada agente.

Colocam-se assim várias questões:

- Quais os fatores que mais influenciam o valor das habitações nos municípios em Portugal?
- Qual o comportamento das oscilações de valor da habitação em Portugal?
- Quais as semelhanças entre municípios no que respeita a estas oscilações?

## 2.2 Análise Bibliométrica

Com o intuito de analisar a importância do tema deste trabalho em termos de investigação científica, optou-se por fazer uma análise bibliométrica sumária. A análise efetuou-se em Janeiro de 2021 através do *Web of Science*, utilizando a pesquisa "housing prices" como tópico, obtendo-se 2369 resultados. Refinando a pesquisa apenas por artigos científicos, resultaram 2003 artigos que se debruçaram sobre esta temática, o que demonstra a crescente importância do tema.

O primeiro artigo publicado sobre o assunto em análise data de 1972. Sensivelmente a partir do ano de 1995 o assunto começou a despertar um interesse maior nos investigadores. O início do século XXI apresentou um crescimento acentuado que se tem verificado até aos dias de hoje, conforme ilustra o gráfico da Figura 2.1

O crescente aumento de interesse pelo tema, poderá estar relacionado com a explosão da bolha imobiliária presente um pouco por todo o mundo, influenciada pelo aumento dos valores

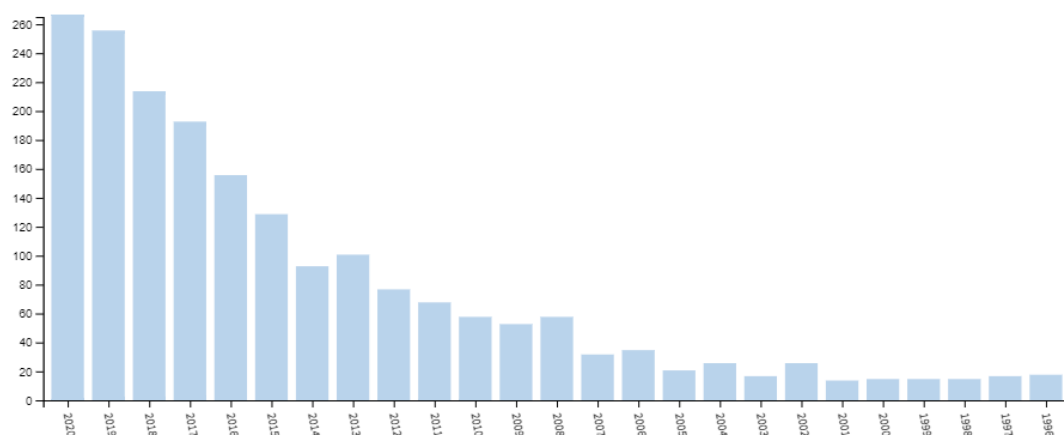


Figura 2.1: Publicações Tema Housing Prices

Fonte: *Web of Science*

dos imóveis, que não foram acompanhados pelo aumento dos rendimentos da população. O tema continua muito atual, tendo inclusive, na data da pesquisa, já 6 artigos para publicação no ano de 2021. À data de hoje (30 de junho de 2021) este número já aumentou para 140 (6.272 %) de todas as publicações nesta temática.

Os dez autores que mais publicaram sobre o assunto foram Hui (20), Tsai (19), Gupta (14), Song e When (12), Glaeser, Su e Zhang (10) e ainda Chau e Quigley(9), tal como se ilustra no gráfico da Figura 2.2.

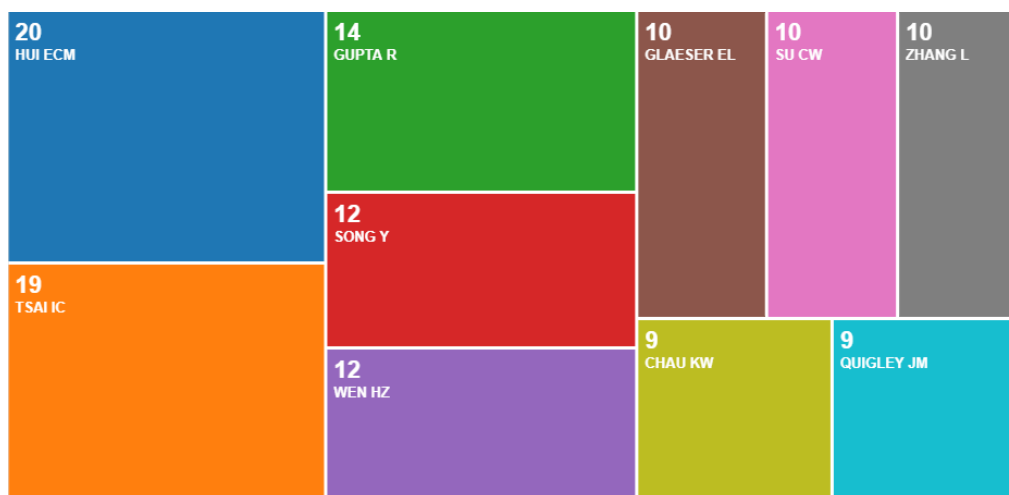


Figura 2.2: Autores com Maior Número de Publicações

Fonte: *Web of Science*

Trata-se de uma área com mais interesse de investigação em países como EUA, República Popular da China, Espanha, Coreia do Sul, Inglaterra, Taiwan, Austrália, Canadá como explanado na Figura 2.3 . De referir que em Portugal existem 14 artigos científicos sobre esta temática, o que denota a importância do tema.

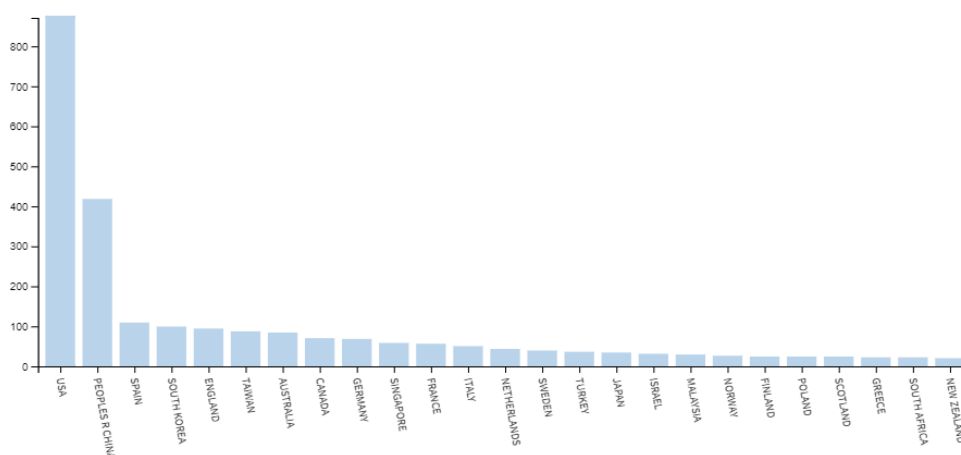


Figura 2.3: Países com Maior N.º de Publicações  
Fonte: *Web of Science*

O tema em análise ganhou especial interesse a partir do ano de 2008, talvez devido à bolha imobiliária dos EUA, a qual levou a um verdadeiro *tsunami* económico e financeiro a nível mundial. Daí a necessidade de estudar as variações de valor das habitações, e quais os fatores que contribuem para essas flutuações.

## 2.3 Fatores Determinantes para o Valor das Habitações Transacionadas

Na literatura são apontados, então, diversos fatores que influenciam os valores das habitações transacionadas. Alguns estudos neste âmbito são de caráter internacional como os de Winston e Eastaway (2008) e de Borio e McGuire (2004), outros mais regional, como Archer, Ling e Smith (2010), Brasington, Hite e Jauregui (2015), Gau e Wang (1994), Guerreiro (2012), Johnes e Hyclak (1999), Liu e Z. Li (2018), S. McGreal e Taltavull de La Paz (2013) e Moreira (2000) ou acerca da proximidade a grandes centros como Alkana et al. (2015), Archer, Ling e Smith (2010), Chasco Yrigoyen e Sánchez Reyes (2012), Fitch Osuna, Soto Canales e Garza Mendiola (2013) e Katz e Rosen (1987), mas a literatura é escassa ao nível de municípios, sendo, desta forma, este trabalho um contributo inovador, também em termos teóricos. Noutra perspetiva a dissertação de Varão 2019 centrou o seu estudo na análise dos factores estruturais e políticos no período de 2000 a 2015.

Nesta secção apresentam-se alguns dos fatores apontados na literatura como determinantes para a definição dos valores das habitações transacionadas, nos âmbitos referidos.

### 2.3.1 Localização

De acordo com Winston e Eastaway (2008) a qualidade de vida depende de vários aspectos, como a localização da casa, como residir numa área limpa e segura, com acesso a recursos naturais, como espaços verdes e/ou água. Segundo estes autores, dada a importância de alguns aspetos da habitação, como a sua localização, design e uso para o desenvolvimento sustentável, a União Europeia e outras organizações internacionais devem adotar indicadores relacionados à habitação nos seus indicadores de desenvolvimento sustentável.

Eddie CM Hui et al. (2007) analisaram um conjunto abrangente de atributos da localização, em contraste com outros estudos que apenas examinam um ou alguns atributos. No seu estudo usaram um modelo hedónico de ajustes espaciais com técnicas GIS (Geographic Information System), como forma de contribuir para a compreensão da importante questão teórica de quanto está disposta a gastar uma família nos vários atributos ambientais e de vizinhança de um apartamento. Os autores abordam a localização através de um método mais apropriado tendo em conta o tempo mais curto para o destino, ou seja, foi tido em consideração o tempo de viagem mais do que a distância física. Os autores concluíram que o tempo de viagem do apartamento para o destino, tem uma correlação negativa com o preço da habitação, e com a disposição das famílias para pagar mais por apartamentos com vista para o mar e localizados em áreas com melhor qualidade do ar.

Kaya e Atan (2014) através de um modelo de preços hedónicos, com dados de venda de habitações na Turquia, durante o mês de Janeiro de 2016, concluíram que um dos fatores mais importantes para a aquisição de habitação é a sua localização, afirmando que, nos dados estudados, os inquiridos estavam dispostos a pagar muito mais por uma residência em Istambul, do que noutros locais. Alkana et al. (2015) analisaram o mercado imobiliário em Ancara, considerando os preços oferecidos nas páginas web em 2011 e uma amostra de 149 observações. Através de uma análise de *clusters*, uma análise discriminante e uma análise de preços hedónicos concluíram que a localização era o fator mais relevante na aquisição de habitação.

S. McGreal e Taltavull de La Paz (2013) analisaram os atributos que foram determinantes na construção dos preços com uma amostra de 2 362 800 preços da habitação em 16 anos (1995–2010), localizadas em sete províncias da Espanha. Através das metodologias STAR (Situation, Task, Action, Results) e GLM (Generalized Linear Model), os resultados destes autores mostram que as características de mercado são específicas para as diferentes regiões.

Assim, tendo em conta os estudos apontados, pode formular-se a seguinte hipótese de investigação:

*H<sub>1</sub>: A localização afeta o valor das habitações.*

### 2.3.2 Densidade Populacional

De acordo com o INE<sup>3</sup>, em 2017 Portugal tinha uma população residente estimada em 10 291 027 pessoas, das quais 4 867 692 homens e 5 423 335 mulheres. Existe uma tendência de decréscimo da população verificada desde 2010, ainda que nos últimos quatro anos se tenha atenuado essa tendência. Além disso, as alterações sofridas ao nível da estrutura etária, denunciam um decréscimo da população jovem e um aumento da população idosa. Verifica-se assim uma tendência para o envelhecimento demográfico, fruto da queda da natalidade e do aumento da esperança média de vida. A idade média da população residente em Portugal aumentou de 42,7 para 44,2 anos, entre 2012 e 2017.

Zhou et al. (2020) efetuaram uma abordagem de dados em painel espacial sobre o efeito dos preços da habitação na produtividade, em 283 cidades chinesas durante o período de 2000–2013. Estes concluíram que os preços da habitação nas cidades chinesas estão positiva e significativamente associados à densidade populacional. Segundo os autores, o aumento

<sup>3</sup>INE – Instituto Nacional de Estatística, disponível em [https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_destaques&DESTAQUESdest\\_boui=315156710&DESTAQUESmodo=2&xlang=pt](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=315156710&DESTAQUESmodo=2&xlang=pt)

dos preços da habitação e a intensificação da população são forças motrizes essenciais para o aumento da produtividade.

O estudo de Rappaport (2008) indica o efeito da densidade populacional no preço das habitações. No seu estudo, efetuado nos EUA, este conclui que vários resultados endógenos co-variam com a densidade populacional. Os salários caem ligeiramente com a densidade populacional, os preços das casas aumentam moderadamente e os preços dos terrenos sobem abruptamente. Este autor conclui que a compensação pela alta qualidade de vida é principalmente capitalizada nos preços da terra e das casas e não nos salários. Em troca de desfrutar de alta qualidade de vida, os indivíduos sacrificam pequenas quantidades de bens comercializados e consumo de lazer e grande quantidade de consumo de habitação. Assim pode formular-se a seguinte hipótese de investigação:

*H<sub>2</sub>: A densidade populacional afeta o valor das habitações.*

### 2.3.3 Poder de compra

Calcagno, Fornero e Rossi (2009) sugerem, no seu estudo sobre o comportamento do consumo das famílias italianas, que o consumo das famílias está positiva e significativamente relacionado com os ganhos de capital das famílias em habitação. Em famílias jovens (onde o chefe da família tem menos de 40 anos) essa relação é positiva, embora apenas marginalmente significativa. Embora as famílias antigas pareçam ser capazes de lucrar com pelo menos uma pequena quota do aumento da riqueza real líquida no seu portfólio imobiliário, esse não é o caso das famílias jovens. Quando os preços das casas aumentam, estes últimos são confrontados com uma expectativa de rendas futuras mais altas.

A acessibilidade dos portugueses ao mercado de habitação tem aumentado, de acordo com a análise da taxa de esforço com habitação no conjunto dos orçamentos das famílias portuguesas. Esta taxa, embora tendendo a agravar-se, é, no entanto, bastante inferior às médias comunitárias.

De acordo com Alves e Pinheiro (2017) a melhoria do mercado de trabalho verificada nos anos precedentes ao ano de 2017 traduziu-se na recuperação do rendimento disponível das famílias, essa expectativa de que no médio-prazo o crescimento se manteria ligeiramente acima de 2% deu suporte à recuperação do rendimento, promovendo a dinamização do mercado imobiliário. S. McGreal e Taltavull de La Paz (2013) destacam a importância do rendimento da população, acessibilidade e características estruturais, para explicar o preço da habitação.

Por outro lado, o aumento dos índices de riqueza é impulsionado pelo aumento dos preços da habitação, bem como alterações nos preços de outros ativos financeiros, segundo Fuller, Johnston e Regan (2020). Os autores ainda salientam que a dinâmica dos preços da habitação e da desigualdade de riqueza moldará cada vez mais as gerações na Europa Ocidental.

Assim, pode ser formulada a seguinte hipótese de investigação:

*H<sub>3</sub>: O poder de compra afeta o valor das habitações.*

### 2.3.4 Dimensão das famílias

Segundo o INE <sup>4</sup>, o agregado doméstico privado diz respeito ao conjunto de pessoas que vivem no mesmo alojamento, em comunhão de casa e mesa, independentemente dos laços de parentesco. A exigência de mobilidade residencial das famílias é definida por diversos fatores, por exemplo, as necessidades que surgem relacionadas com as alterações da dimensão do agregado familiar.

De acordo com Varão (2019) os impactos demográficos demonstraram ser os mais relevantes para a variação dos valores das habitações, já que um aumento da população, seja através de um aumento da imigração ou das taxas de natalidade, afetam o preço das casas positivamente. Este aumento é explicado por um consequente aumento da procura de habitação (afetando os preços), especialmente se a oferta demorar muito tempo a ajustar (Varão 2019). No estudo de Archer, Ling e Smith (2010), efetuado no mercado habitacional de Chicago, os autores descobriram que uma das características das famílias que têm o poder dominante para explicar a variação transversal dos preços das habitações é o tamanho da família. Assim argumentam que as mudanças neste tamanho também influenciam a duração da permanência das famílias numa dada propriedade.

Assim, pode ser formulada a seguinte hipótese de investigação:

*H<sub>4</sub>: A dimensão das famílias afeta o valor das habitações.*

### 2.3.5 Desemprego

Johnes e Hyclak (1999) efetuaram um estudo sobre a existência de potenciais interações entre a habitação local e os mercados de trabalho. Utilizaram um modelo que relaciona o salário médio da indústria, a taxa de desemprego, a força do trabalho e o preço médio de uma casa numa zona urbana. As estimativas de regressão foram calculadas para as áreas de Hartford, Houston, Fort Lauderdale e Milwaukee, usando dados trimestrais para os anos 80. Encontraram algumas evidências de que o desemprego e a força de trabalho podem afetar os preços das casas, e que os preços das casas têm um efeito significativo sobre o tamanho da força de trabalho.

Segundo Robson (2003), no início dos anos 90, a queda acentuada no mercado imobiliário no sudeste da Inglaterra, resultou num declínio nos preços das casas regionais de mais de sete por cento, e contribuíram para um aumento acentuado na taxa de desemprego na região, tornando mais difícil a manutenção dos postos de trabalho nas empresas. A taxa de saída dos indivíduos da situação de desemprego também sofreu uma redução, contribuindo para o aumento acentuado da taxa regional de desemprego. Em contrapartida, nas regiões mais tradicionais de elevado desemprego, como o Norte e Escócia, a perturbação para o mercado imobiliário, e, portanto, o efeito sobre as entradas e saídas do desemprego, foram muito menos graves.

Borio e McGuire (2004), no estudo efetuado entre as décadas de 1970 a 2000, compreendendo 13 países industrializados: Austrália, Canadá, Dinamarca, Finlândia, Irlanda, Japão, Holanda, Noruega, Espanha, Suécia, Suíça, Reino Unido e EUA, defendem que os picos dos preços das ações anunciavam os picos subsequentes dos preços da habitação e identificam como determinantes tradicionais dos preços da habitação as taxas de juros, o crescimento da produção e o desemprego.

<sup>4</sup><https://smi.ine.pt/Conceito/Detalhes/5498>

Podendo então, com base na literatura apresentada, formular-se a seguinte hipótese de investigação:

*H<sub>5</sub>: O desemprego afeta o valor das habitações.*

### 2.3.6 Qualidade Ambiental

Xiao, Eddie C.M. Hui e Wen (2020) no seu estudo utilizaram um modelo de preços hedónico e o modelo econométrico espacial para investigar os efeitos externos da quadrilha no mercado imobiliário de Hangzhou, China. A quadrilha é uma atividade pública popular realizada em praças e parques das cidades chinesas, onde os residentes mais próximos podem sofrer com a poluição sonora e outras externalidades negativas. Os resultados mostram que o preço da habitação nas proximidades caíram 5,8% e 13,0% quando as pessoas dançam no parque e na praça mais próximos. Os resultados deste estudo indicam que a quadrilha tem um efeito externo negativo significativo no preço da habitação e no valor do espaço aberto.

Em Machala (Equador), Zambrano-Monserrate (2016) analisou o mercado imobiliário de apartamentos para arrendamento em 2013, com uma amostra de 635 observações. As características que se mostraram mais importantes para a aquisição foram as que se referem à existência de serviços públicos, como abastecimento de água e a recolha de lixo.

Chasco Yrigoyen e Sánchez Reyes (2012) analisaram como a poluição do ar e o ruído afetam os preços das casas, no centro de Madrid, por meio de uma regressão quantílica. Como determinantes consideraram os atributos estruturais da moradia, variáveis de acessibilidade e características sociais e ambientais. Houve regressão espacial do quantil com nove deciles, onde os coeficientes autoregressivos foram muito significativos, especialmente nos dois últimos, correspondendo às residências com preços mais altos. Por outro lado, moradias mais acessíveis (deciles 0,1 e 0,2), localizadas em áreas não tão centrais, podem ter um problema de autocorrelação espacial, o que se torna evidente no menor poder explicativo do modelo. Em todos os deciles, a poluição do ar tem um efeito negativo no preço, mas mais significativamente nas propriedades mais caras. Por outro lado, o ruído não foi significativo, exceto para casas de luxo ou do último decil localizadas no centro de Madrid, onde o aumento do ruído acarreta um aumento no preço.

Fitch Osuna, Soto Canales e Garza Mendiola (2013) debruçaram o seu estudo no município de San Nicolás de los Garza, no México, uma área imersa numa dinâmica metropolitana. Consequentemente, o processo imobiliário é avaliado no contexto das forças centrais e periféricas que geram o crescimento e o ciclo de vida urbano e habitacional, bem como a qualidade do ambiente urbano. Utilizaram um modelo de preços hedónico para definir os atributos que explicam, com maior grau de significância, a variação do valor do terreno. Obtiveram resultados semelhantes aos de Chasco Yrigoyen e Sánchez Reyes (2012), ou seja, os preços da habitação aumentam em áreas mais barulhentas, pois a localização e as características demográficas prevalecem sobre outros fatores.

Assim, pode ser formulada a seguinte hipótese de investigação:

*H<sub>6</sub>: A qualidade ambiental afeta o valor das habitações.*

### 2.3.7 Envelhecimento

Ermisch (1996) faz uma abordagem micro econométrica sobre qual o efeito das mudanças na distribuição etária da população na Grã-Bretanha, na procura por habitação e os preços

de casas. Ele estimou uma função de procura por habitação e, em seguida, examinou as implicações das mudanças na distribuição de idade na procura por habitação. Concluiu que existe uma desaceleração substancial no crescimento da taxa da procura, decorrente de mudanças de distribuição etária, como consequência do envelhecimento da população.

De acordo com Bayet et al. (1991), a despesa que é colocada em primeiro lugar na estrutura do orçamento familiar, é a despesa relacionada com a habitação, sendo que os valores percentuais mais elevados correspondem aos grupos etários: jovens (- 30 anos) e idosos (65 e + anos).

Eichholtz e Lindenthal (2014) no seu estudo sobre as famílias inglesas em 2001, afirmam que altos níveis de educação, boa saúde e altos rendimentos aumentarão a procura das famílias por moradia, mesmo quando estas envelhecem. Trata-se portanto, na sua perspectiva, de um bom sinal para a mercado imobiliário europeu com uma população fortemente envelhecida.

Pode então formular-se a seguinte hipótese de investigação:

*H<sub>7</sub>: O envelhecimento afeta o valor das habitações.*

### **2.3.8 Nível de Educação**

Alguns autores afirmam que o nível de educação das famílias tem efeito na compra de habitação. De acordo com Brasington, Hite e Jauregui (2015) no seu estudo realizado em Ohio, EUA, o salário e a discriminação educacional reduzem os valores de aquisição da habitação.

Liu e Z. Li (2018), com base em dados recolhidos em Chengdu, durante o ano de 2016, para um grupo de pessoas com nível de formação elevado, e com recurso a inquéritos efetuados por via digital, encontraram alguns resultados interessantes. Os entrevistados procuravam independência e um sentimento (individual) de pertença na compra da sua primeira casa, enquanto que para a compra de uma segunda ou mais casas, as forças motrizes mais significativas para a aquisição foram: ficar com a casa para morar com a família ou para investimento. Esta pesquisa forneceu uma nova compreensão sobre o comportamento de compra de habitação dos cidadãos chineses urbanos, especialmente aqueles que têm um maior nível de formação.

No seu artigo sobre a procura por residência, Eichholtz e Lindenthal (2014) concluíram que o mercado imobiliário depende da idade e de outras características demográficas ao nível do agregado familiar. Com base numa amostra transversal de famílias inglesas em 2001, descobriram que o capital humano, representado por educação e saúde, é o principal fator para a procura por habitação. Por exemplo, por cada nível profissional de educação adicional que a família atingiu, a sua disposição para pagar por serviços de habitação é maior.

Gyourko e Linneman (1996) afirmam que embora fatores demográficos, como estado civil e estrutura familiar continuem a influenciar a opção pela posse de habitação, o seu impacto diminuiu, especialmente entre as famílias com melhor nível educacional e as com rendimento real crescente. Segundo os autores, as condições do mercado de trabalho, conforme evidenciado pelos retornos crescentes da qualificação, são sentidas com mais força do que nunca no mercado imobiliário.

Pode assim considerar-se a seguinte hipótese de investigação:

*H<sub>8</sub>: O nível de educação afeta o valor das habitações.*

### 2.3.9 Estado Civil

Yanan (2017) através do seu artigo explorou as relações da estrutura demográfica e os preços da habitação em Macau. Utilizando o modelo de regressão, analisou a estrutura demográfica e o preço das habitações de Macau de 2005 a 2015. Encontrou evidências de que o nível de envelhecimento, o nível de dependência infantil, o estado civil e o PIB desempenham papéis significativos no preço da habitação em Macau.

Huang e Clark (2002) usaram uma pesquisa nacional de 1996 sobre a habitação na China e uma técnica de modelação multinível, para estudar a escolha do direito à habitação na China urbana, onde às famílias é concedida uma liberdade limitada de escolha no mercado imobiliário. O estudo empírico expõe que tanto os fatores socio-económicos como os institucionais afetam as escolhas de compra das famílias na China urbana. O estudo conclui que as famílias mais velhas, de maior dimensão e com maior rendimento disponível, têm uma propensão maior para a compra da habitação. Por outro lado, pessoas casadas e famílias com mais trabalhadores são menos propensos a comprar, o que pode estar relacionado com a deslocação entre domicílio e as diferentes unidades de trabalho.

Já há muitos anos que Chevan (1982) se concentrou na relação entre fatores demográficos e status de compra, ou seja, idade, sexo, estado civil, educação, entre outros, na tomada de decisão para aquisição de habitação. No seu estudo, com dados da fonte estatística *US Census Bureau*, concluiu que, geralmente, uma família se muda para uma casa grande com a intenção de permanecer lá, pelo menos até que o tamanho da família comece a diminuir.

Para Gyourko e Linneman (1996) os fatores demográficos, incluindo estado civil e tipo de família, ainda são muito importantes na decisão de escolha da compra de habitação, mas a sua influência está a diminuir, particularmente entre as famílias mais instruídas com rendimento real crescente.

Assim pode considerar-se a seguinte hipótese de investigação:

*H<sub>9</sub>: Os fatores demográficos afetam o valor das habitações.*

### 2.3.10 Habitação Própria

Tsai e Peng (2011) testaram a existência de especulação (bolha) no mercado imobiliário de Taiwan. Através de dados em painel estudaram a relação do mercado com a taxa de hipotecas, a oferta de moeda, a taxa de inflação, a taxa de crescimento económico, a taxa de casa própria e o preço da habitação para o proprietário. Os autores relacionaram a taxa de casa própria e o indicador de especulação (bolha) e concluíram que a especulação dificulta a propriedade de casa própria, embora a taxa de casa própria aumente anualmente em Taiwan, porque os chineses preferem possuir imóveis do que arrendar. Concluíram ainda que o custo do proprietário na habitação está altamente relacionado aos preços de habitação, indicando que a especulação imobiliária é um fardo pesado para a população. Para cumprir com os custos da habitação, os proprietários precisam de reduzir o consumo de outros bens, o que se traduz numa perda na qualidade de vida. Desta forma Tsai e Peng (2011) acreditam que as flutuações (subida ou descida) de preços das casas têm implicações políticas importantes. De acordo com as suas conclusões o governo chinês não pode simplesmente interferir nos preços baixos das casas enquanto negligencia a especulação imobiliária (bolha).

Gau e Wang (1994), tendo por base dados dos apartamentos e mercado comercial de Vancouver, Canadá, observando mais de 1000 transações de aproximadamente 750 propriedades que foram vendidas no período 1971-1985, realizaram vários testes empíricos e encontraram

períodos médios de detenção de cerca de cinco (apartamento) e oito anos (comerciais) e uma distribuição de períodos de detenção que se assemelha a uma função exponencial negativa. Concluíram que a decisão do período de detenção não é baseada em benefícios fiscais, mas é sim num reflexo das condições de mercado no momento da venda e da consideração de riqueza do investidor.

Archer, Ling e Smith (2010) no seu estudo contendo dados das moradias transacionadas no estado da Califórnia, concluíram que as taxas de rotatividade são significativamente afetadas pelo tipo de posse na venda. No entanto, o efeito reverte-se ao longo do tempo em que as casas para alugar apresentam taxas de rotatividade mais altas nos primeiros oito anos de propriedade, mas taxas de rotatividade mais baixas a partir de então.

Pode assim considerar-se a seguinte hipótese de investigação:

*H<sub>10</sub>: O tipo de propriedade afeta o valor das habitações.*

Fatores/Determinantes	Hipóteses	Sinal Esperado	Autor
Localização	Hipótese 1	(+/-)	Winston e Eastaway (2008) Eddie CM Hui et al. (2007) Kaya e Atan (2014) Alkana et al. (2015) S. McGreal e Taltavull de La Paz (2013)
Densidade Populacional	Hipótese 2	(+)	Rappaport (2008) Zhou et al. (2020)
Poder de Compra	Hipótese 3	(+)	Calcagno, Fornero e Rossi (2009) S. McGreal e Taltavull de La Paz (2013) Fuller, Johnston e Regan (2020)
Dimensão das Famílias	Hipótese 4	(+)	Varão (2019) Archer, Ling e Smith (2010)
Desemprego	Hipótese 5	(-)	Johnes e Hyclak (1999) Robson (2003) Borio e McGuire (2004)
Qualidade Ambiental	Hipótese 6	(+/-)	Xiao, Eddie C.M. Hui e Wen (2020) Zambrano-Monserrate (2016) Chasco Yrigoyen e Sánchez Reyes (2012) Fitch Osuna, Soto Canales e Garza Mendiola (2013)
Envelhecimento	Hipótese 7	(+/-)	Ermisch (1996) Bayet et al. (1991) Eichholtz e Lindenthal (2014)
Nível de Educação	Hipótese 8	(+/-)	Brasington, Hite e Jauregui (2015) Liu e Z. Li (2018) Eichholtz e Lindenthal (2014) Gyourko e Linneman (1996)
Estado Civil	Hipótese 9	(+/-)	Yanan (2017) Huang e Clark (2002) Chevan (1982) Gyourko e Linneman (1996)
Habitação Própria	Hipótese 10	(+/-)	Tsai e Peng (2011) Gau e Wang (1994) Archer, Ling e Smith (2010)
Habitações existentes	Hipótese 11	(+/-)	Xu (2017) Katz e Rosen (1987)

Tabela 2.1: Fatores, sinal esperado e autores

### 2.3.11 Habitações existentes

No seu estudo efetuado na China Xu (2017) refere que se a oferta de habitação está na sua capacidade máxima (ou perto dela) existe uma maior pressão sobre a oferta, o que pode de

facto aumentar os preços das habitações.

Katz e Rosen (1987) mostrou empiricamente, baseado num grande conjunto de dados sobre habitação na área da Baía de São Francisco, na Califórnia, que as taxas aplicadas na construção parecem ter tido um efeito substancial nos preços das habitações. Os resultados obtidos, através de um modelo de regressão, indicam que os preços das casas estão entre 17 por cento e 38 por cento mais altos nas comunidades em que a moratória de crescimento e/ou planos de controle de crescimento estão presentes. O uso generalizado de taxas, em muitas comunidades, limitam a resposta de oferta disponível. A disseminação dessas técnicas reguladoras do crescimento imobiliário para áreas metropolitanas fora da Califórnia podem ter efeitos negativos substanciais na acessibilidade de moradia (Katz e Rosen 1987).

Assim, pode ser considerada a seguinte hipótese de investigação:

*H<sub>11</sub>: A número de habitações disponíveis afeta o valor das habitações.*

### 2.3.12 Síntese da Revisão de Literatura

Na Tabela 2.1 apresenta-se um resumo dos determinantes apontados na literatura para a variação do valor das habitações, a correspondente hipótese de investigação, o sinal esperado do impacto desse determinante/fator e alguns dos autores que o indicaram como significativo.

## 2.4 Métodos Quantitativos de Apoio à Decisão

Tendo em conta os objetivos deste trabalho, definidos na Secção 1.2, será necessário utilizar diversos métodos quantitativos de apoio à decisão, em particular, a Regressão Linear Múltipla (Secção 2.4.1) e a Análise de Clusters (Secção 2.4.2).

Assim, nesta secção será apresentada uma breve descrição destas metodologias. Para concretização do objetivo (O2) serão consideradas técnicas descritivas.

### 2.4.1 Regressão linear

Marôco (2010) refere que o termo “Análise de Regressão” determina um vasto conjunto de técnicas estatísticas usadas para modelar relações entre variáveis e predizer o valor de uma ou mais variáveis dependentes (ou de resposta) a partir de um conjunto de variáveis independentes (ou predictoras).

Hair et al. (2010) refere que a regressão linear múltipla é uma técnica multivariada usada para estabelecer relações entre variáveis e predizer o valor de uma variável dependente ou explicada (métrica) a partir de um conjunto de variáveis independentes ou explanatórias (métricas/não métricas). Nesta técnica é usual utilizar o método dos mínimos quadrados (OLS – *ordinal least squares*) para estimar a contribuição de cada variável independente para explicar a variável dependente.

As técnicas inferenciais são componentes importantes na análise de dados. No entanto quando se constata a presença de heterocedasticidade, toda a análise baseada em testes de hipóteses se torna inválida. Existe então a necessidade de ajustar os erros-padrão e as estatísticas t e F, obtidos por mínimos quadrados. Perante o caso em que a matriz é conhecida, existe a possibilidade de estimar os parâmetros do modelo de regressão a partir de um método chamado de Mínimos Quadrados Generalizados (GLS). O objetivo do método é estimar os parâmetros do modelo, tal como na metodologia OLS, no entanto é levada em

consideração a nova configuração da matriz de variâncias e covariâncias dos erros, ou seja, o modelo é transformado, de forma a obter um termo de erro homocedástico.

Assim, no modelo de regressão linear, a variável dependente ou resposta  $Y$  pode estar relacionada com  $k$  variáveis explicativas ou independentes, ou seja, a variável  $Y$  é modelada como função linear de vetores multidimensionais, onde o número de atributos preditores é variável (Hair et al. 2010).

O modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u_i \quad (2.1)$$

recebe o nome de regressão linear múltipla com  $k$  variáveis explicativas e os parâmetros  $\beta_j, j = 0, \dots, k$  designam-se por coeficientes de regressão, sendo que  $u_i, i = 0, \dots, k$ , correspondem ao erros aleatórios. Assim se há uma única variável preditora, a função descreve uma linha reta. Se houver duas variáveis predictoras, então a função descreve um gráfico no plano. Se existem  $n$  variáveis predictoras, então a função descreve um hiperplano  $n$ -dimensional.

Para estimação dos coeficientes de regressão, usa-se o método dos mínimos quadrados, devendo proceder-se à análise do resultado obtido para o teste global, cujas hipóteses são:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_n = 0 \text{ (o modelo não é significativo)}$$

vs

H1: Existe pelo menos 1 coeficiente de regressão que é significativamente diferente de zero (o modelo é significativo).

Realizado o teste global, quando este apresenta um  $p\text{-value} < \alpha\%$  (usualmente  $\alpha = 5$ ) então para um nível de significância de 5% rejeita-se a hipótese nula, logo pode concluir-se que existe pelo menos 1 coeficiente que é significativamente diferente de zero, ou seja, existe pelo menos uma variável independente que contribui para explicar a variável dependente e, conseqüentemente, o modelo de regressão linear estimado é significativo.

Posteriormente, procede-se à análise aos testes marginais para cada variável independente  $X_j, j = 0, \dots, n$ , cujas hipóteses são:

$$H_0: \beta_j = 0 \text{ vs } H_1: \beta_j \neq 0, j = 0, \dots, n.$$

Na presença de um  $p\text{-value} < 5\%$ , rejeita-se a hipótese nula para um nível de significância de 5%, ou seja, o coeficiente  $b_j$  é significativamente diferente de zero, o que significa que a variância da variável  $X_j$  contribui para explicar a variância da variável  $Y$ .

No sentido de analisar a qualidade do modelo de regressão linear pode analisar-se o coeficiente de determinação ajustado (*Adjusted R-squared* – R-quadrado ajustado –  $R_a^2$ ), ou seja, o poder explicativo do modelo. O coeficiente de determinação é a percentagem da variação da variável resposta que é explicada por um modelo linear. Este coeficiente tem um valor entre 0 e 1 (0 e 100%), em que 0% indica que o modelo não explica nada da variabilidade (em torno da sua média) da variável resposta e 100% indica que o modelo explica toda a variabilidade da variável resposta.

Em geral, quanto maior o R-quadrado, melhor o modelo se ajusta aos dados. No entanto, é necessário que se verifiquem todas os pressupostos do modelo de regressão, que são:

- $Y_i$  é uma variável aleatória (pode estar afetada por erros de medição);

- $X_i$  são medidas sem erro ou, quanto muito, esse erro é desprezível face à variação das variáveis preditoras  $X_j$ ;
- $X_i$  são ortogonais (i.é. não estão correlacionadas ou, quanto muito, apresentam correlações fracas) – ausência de multicolinearidade;
- $u_i$  são variáveis aleatórias, independentes e seguem distribuição normal de média nula e variância constante (Análise de Resíduos).

### Multicolinearidade

A multicolinearidade diz respeito à correlação entre duas variáveis explicativas ou entre uma delas e as demais incluídas na equação do modelo de regressão. A multicolinearidade ocorre quando, duas variáveis medem aproximadamente o mesmo, ou seja, a correlação entre elas é alta. Como forma de detectar relações colineares podem utilizar-se diversas medidas: coeficiente de correlação amostral, a tolerância, o VIF, os valores próprios e o *Condition Index*.

O coeficiente de correlação amostral entre os pares de variáveis explicativas. Frequentemente, um coeficiente de correlação entre duas variáveis explicativas, superior a 0,80 ou 0,90 em valor absoluto denota uma forte associação linear e uma relação de colinearidade inadequada ao modelo de regressão.

Para verificação da ausência ou não de multicolinearidade, pode ainda ser utilizada a medida  $T$  (Tolerância), que mede o grau em que uma variável independente é explicada pelas variáveis independentes, sendo que  $T = 1 - R^2$ , onde  $R^2$  é o coeficiente de determinação entre a variável em questão e as outras variáveis independentes. O  $T$  varia entre 0 e 1, sendo que a proximidade de zero denota a existência de multicolinearidade. O valor limite é 0,1, logo as variáveis com valores de tolerância inferiores a 0,1 devem ser excluídas do modelo, segundo Hair et al. (2010).

Outra análise a efetuar é ao valor do VIF (*Variance Inflation Factor*) das variáveis preditoras. Existem diversas abordagens relativamente aos limites para este indicador. Hair et al. (2010) considera a ausência de multicolinearidade se VIF for inferior a 3. Outros autores indicam que há casos em que se pode admitir que os valores de VIF superiores a 5 (ou 10) demonstram a existência de multicolinearidade (Marôco 2010).

Se a análise aos valores próprios (*Eigenvalue*) demonstrar valores próximos de 0 então estamos perante a existência de multicolinearidade (Hair et al. 2010; Marôco 2010).

Analisando os *Condition Index* ( $k'$ ), se estes forem superiores a 15, então estamos na presença da existência de possíveis problemas, quando os  $k'$  se apresentam superiores a 30, estamos perante a existência de multicolinearidade (Marôco 2010).

Na presença de coeficientes estimados afetados de forma significativa pela multicolinearidade, existe a necessidade de encontrar uma solução mais adequada. Este problema não é de solução fácil. Alguns dos procedimentos possíveis são: aumentar o tamanho da amostra ou exclusão das variáveis colineares.

### Análise de Resíduos

A análise de resíduos investiga as características que comprometem a validade do modelo de regressão linear. Há 3 aspetos a considerar: os resíduos devem ser independentes, seguir uma distribuição normal de média nula e variância constante (homogeneidade).

Primeiramente deve verificar-se a condição  $Cov(u_k, u_l) = 0$ , com  $k$  e  $l \in 1, \dots, k$ , ou seja os erros são independentes. Para tal efetua-se o teste de *Durbin-Watson*. Se for verificado um valor próximo de 2 na estatística de *Durbin-Watson* então está-se perante erros não correlacionados.

De seguida deve verificar-se a condição  $u_i \sim N(0, \Sigma)$ , ou seja, se os erros possuem distribuição normal de média nula e variância constante. A verificação pode ser feita através da análise do diagrama de dispersão ou *Scatterplot*, do *Pp – plot*/histograma, e através de testes de normalidade, por exemplo o Teste de *Kolmogorov-Smirnov* ( $K - S$ ) ou de *Sahpiro-Wilk* (*SW*).

### 2.4.2 Análise de Clusters

De acordo com Hair et al. (2010) a análise de *clusters* é um conjunto de procedimentos estatísticos multivariados, usados para classificar elementos em grupos, de forma que elementos dentro de um mesmo *cluster* sejam muito parecidos, e os elementos em diferentes *clusters* sejam distintos entre si.

Para definir as semelhanças, ou dissimilaridades entre os elementos é usada uma função de distância, que deve ser definida considerando o contexto do problema em questão. Uma dificuldade inicial na análise de *clusters* é que não há um único critério, similaridade, medida e/ou técnica de definição dos grupos. A literatura sobre o assunto, bem como os pacotes estatísticos disponíveis, apresentam-nos uma vasta gama de critérios, sempre com o objetivo de obter grupos homogéneos e significativamente diferentes uns dos outros (Costa e Silva, Correia e Borges 2021).

De acordo com Costa e Silva, Correia e Borges 2021; Hair et al. 2010 existem dois métodos para formar *clusters*:

- Análise de *clusters* hierárquica – aplica-se tanto a casos como a variáveis;
- Análise não hierárquica (por exemplo o conhecido método *k – means*) – apenas se aplica a casos.

A aglomeração hierárquica caracteriza-se pelo criação de uma hierarquia ou estrutura em forma de árvore. No agrupamento hierárquico de *clusters*, estes formam-se com base nos pares de casos mais próximos com base na medida de distância escolhida. O algoritmo efetua esta tarefa passo a passo (juntando pares de casos, pares de *clusters*, ou um objecto a um *cluster* até existir apenas um *cluster*). O processo é apresentado no Dendrograma. Já na aglomeração não hierárquica, são agrupados casos num conjunto de *clusters* cujo número é definido à partida, mas os casos podem ser realocados caso essa mudança melhore a otimização da similaridade de casos dentro do *cluster* e a dissimilaridade entre os *clusters*.

De acordo com Hair et al. 2010, as particularidades de cada objeto são combinadas com uma medida de semelhança, que pode ser de similaridade ou dissimilaridade, calculada para todos os pares de objetos, permitindo a comparação de qualquer objeto com outro pela medida de similaridade e a associação dos objetos semelhantes por meio da análise de agrupamento.

As medidas de distância representam a similaridade, que é caracterizada pela proximidade entre as observações, tendo em conta o valor das variáveis consideradas. A distância euclidiana é a medida de distância mais frequentemente utilizada quando todas as variáveis são quantitativas.

Cinco etapas devem ser consideradas na aplicação da análise de *clusters*, segundo Reis (2001), apresentadas na Tabela 2.2.

1. <sup>a</sup> Seleção dos dados a agrupar
2. <sup>a</sup> Definir um conjunto de variáveis a partir das quais será obtida informação para agrupar os casos
3. <sup>a</sup> Selecionar a medida de semelhança ou distância entre cada par de casos
4. <sup>a</sup> Escolha do critério de agregação ou desagregação dos dados
5. <sup>a</sup> Validação dos resultados encontrados

Tabela 2.2: Etapas Análise Clusters  
Fonte: Reis (2001)

Para os procedimentos de aglomeração são vários os métodos hierárquicos de agrupamento possíveis, de entre eles o método da distância entre grupos (*Between – groups linkage*), que trata a distância entre grupos como sendo a distância média entre os pares de indivíduos dos dois grupos, tentando agrupar os agregados cuja distância média é menor, o método da menor distância (*Nearest neighbor*) onde a distância entre dois objectos é representada pela distância mínima entre todas as possibilidades de pares de objectos em dois grupos, o método da maior distância (*Furthest neighbor*) que se baseia na distância máxima, ao contrário do método da ligação individual. Neste método a distância entre dois grupos é definida como a distância máxima entre todos os pares de possibilidades de observações nos dois grupos. Procura agrupar elementos cuja distância entre os mais afastados seja a menor. O método do centróide (*Centroid clustering*) baseia-se na distância, normalmente euclidiana ou quadrática euclidiana, entre centróides, privilegiando a menor distância entre eles. Identifica os dois grupos separados pela menor distância entre os pontos mais próximos e coloca-os no mesmo agrupamento. O método de *Ward* (*Ward's method*) que procura agrupar os agregados que apresentam menor soma dos quadrados entre os dois agrupamentos, calculada sobre todas as variáveis (Costa e Silva, Correia e Borges 2021).

Como constatado anteriormente, no método hierárquico, determina-se uma relação de hierarquia entre os sujeitos e os grupos, o que não acontece num método não hierárquico, dado que após a determinação do número de grupos, o processo torna-se dinâmico e interactivo, com o objectivo de identificar a melhor solução. Os métodos não hierárquicos utilizam-se para agrupar indivíduos (e não variáveis), cujo número inicial de *clusters* é definido pelo investigador. Utilizando os métodos não hierárquicos a probabilidade de ocorrerem classificações erradas nos agrupamentos é menor, no entanto existe a dificuldade de se estabelecer o número de *clusters* à partida. De acordo com Marôco (2010) a alternativa consiste na utilização do método hierárquico como técnica exploratória para, posteriormente, propiciar a utilização do número indicado de *clusters* na técnica não hierárquica. O método *k – Means* é o mais conhecido dos métodos não hierárquicos.

Segundo Marôco (2010), o processo no método *k-means* é composto por três passos:

1. Partição inicial dos indivíduos em *k clusters* definidos pelo analista;
2. Cálculo dos centróides para cada um dos *k clusters* e cálculo da distância euclidiana dos centróides a cada sujeito na base de dados;
3. Agrupar os sujeitos aos *clusters* cujos centróides se encontram mais próximos, e voltar ao passo 2 até que não ocorra variação significativa na distância mínima de cada sujeito da base de dados a cada um dos centróides do *k clusters* ( ou até que o número máximo de iterações ou o critério de convergência - definido pelo analista - seja alcançado).

Não há uma forma ideal de escolher o número de *clusters* a considerar, segundo Costa e Silva, Correia e Borges (2021). Podem ser usadas regras como *conhecimento a priori*, a regra de ouro (*rule of thumb*) ou o método do cotovelo, mas o ideal é que os *clusters* resultantes tenham boas propriedades estatísticas, como:

- testes ANOVA significativos;
- convergência do algoritmo;
- *clusters* com número homogêneo de casos .
- medidas de validação adequadas, por exemplo:
  - correlação cofonética – quanto mais próximo de 1 for melhor. Só é aceitável a análise se for maior do que 0,5.
  - medidas de validação interna:
    - \* Conectividade – tem um valor entre 0 e  $\infty$  e deve ser minimizada;
    - \* *Sillhouette Width* – observações bem classificadas no cluster têm valores próximos de 1 e mal classificadas tem valor próximo de -1, então esta medida tem valores no intervalo  $[-1, 1]$  e deve ser maximizada;
    - \* *Dunn Index* – tem um valor entre 0 e  $\infty$  e deve ser minimizado;
  - medidas de validação de estabilidade:
    - \* *Average Proportion of Non-overlap* (APN) – está no intervalo  $[0, 1]$ , com valores próximos de zero correspondentes a resultados de cluster altamente consistentes;
    - \* *Average Distance*(AD) – tem valores entre 0 e  $\infty$  e são preferenciais valores pequenos;
    - \* *Average Distance between Means* (ADM) – tem valores entre 0 e  $\infty$  e são preferenciais valores pequenos;
    - \* *Figure of Merit* (FOM) – tem valores entre 0 e  $\infty$  e são preferenciais valores pequenos;
  - medidas de validação biológica:
    - \* *Biological Homogeneity Index* (BHI) – está no intervalo  $[0, 1]$ , com valores próximos de 1 correspondentes a maior homogeneidade;
    - \* *Biological Stability Index* (BSI) – está no intervalo  $[0, 1]$ , com valores próximos de 1 correspondentes a maior estabilidade.

A validação dos resultados pode ser feita usando o package *cValid* do software *R* ou através da análise discriminante ou através da análise da convergência do algoritmo e de testes de comparação de médias.

## Capítulo 3

# Metodologia e Resultados

### 3.1 Metodologia

Tendo em vista o objetivo principal deste trabalho, que é determinar que fatores influenciam o valor das habitações nos municípios em Portugal, foi efetuada a revisão de literatura apresentada no Capítulo 2, dando maior enfoque aos fatores apontados como os mais significativos.

Depois de efetuada a revisão de literatura e de identificados os fatores apontados como determinantes para a definição do valor das habitações pelos investigadores, este trabalho analisa se estes mesmos fatores se evidenciam como importantes para o objeto de estudo deste trabalho (os municípios), em particular os municípios portugueses. Não obstante os trabalhos dos autores mencionados na Revisão de Literatura, não pretenderem a mesma análise intra-municipal, os seus resultados são muito úteis, podendo ser utilizados como uma base confiável, pois resultaram de pesquisas no mesmo âmbito temático, só que debruçadas sobre âmbitos geográficos diferentes. Este âmbito geográfico deste estudo é o maior diferenciador deste trabalho, dos já existentes.

Neste estudo utiliza-se a regressão linear múltipla para analisar o impacto no valor da habitação em Portugal, dos fatores identificados na literatura, apresentados no Capítulo 2, e resumidos na Tabela 2.1.

Trata-se assim de um estudo econométrico com o objectivo de avaliar qual o impacto de diversas características dos municípios (variáveis independentes) no valor médio das transações de prédios praticado nestes municípios (variável dependente). Pretende-se compreender de que forma a variação destas características municipais influencia a evolução dos valores.

Ainda se pretende estudar o comportamento das oscilações de valores da habitação em Portugal e determinar as semelhanças entre municípios no que respeita a estas oscilações. Para tal consideram-se as técnicas de análise de *clusters* hierárquica e não hierárquica, no sentido de identificar grupos homogêneos e estudar que características os diferenciam.

Estas técnicas foram aplicadas através do software IBM SPSS, versão 25. Também foi usado o software R, versão 4.0.3, em particular para a validação da análise de *clusters*.

### 3.2 Evolução do VMPT em Portugal

Um dos objetivos deste trabalho (O2) é estudar o comportamento das oscilações do valor da habitação em Portugal, para tal consideraram-se os dados anuais disponíveis no PORDATA relativos ao VMPT em Portugal durante o período de 1992 a 2018.

Estes dados incluem os prédios transaccionados: total, urbanos e rústicos e estão ilustrados no gráfico da Figura 3.1.

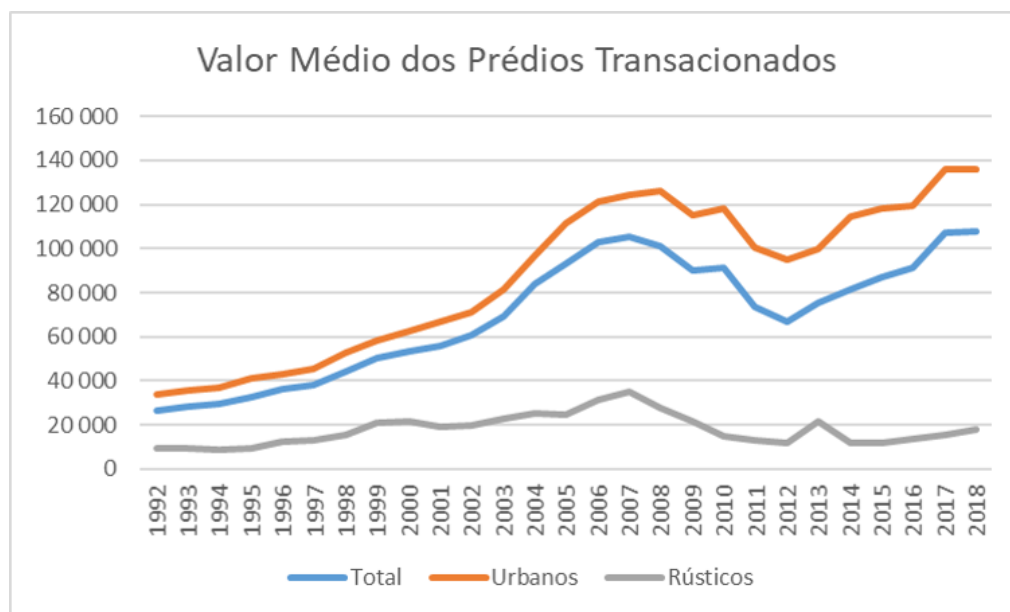


Figura 3.1: Evolução do VMPT em Portugal

Pode observar-se na Figura 3.1. que a evolução do VMPT em Portugal para prédios transaccionados total, urbanos e rústicos é semelhante, sendo sempre o valor médio para prédios urbanos claramente superior ao VMPT dos prédios rústicos.

Os VMPT tiveram uma tendência crescente até 2007, tendo decrescido até 2012, ano a partir do qual voltaram a aumentar.

O período compreendido entre os anos de 2007 e 2012 representa os anos da crise mundial iniciada pela falência do gigante da indústria financeira *Lehman Brothers*, decretada pelo tribunal federal norte-americano no dia 15 de setembro de 2008. Uma decisão que mudaria a história económico-financeira mundial. Portugal por arrasto iniciou um dos seus períodos "negros", tendo inclusive recorrido a assistência financeira externa. A 5 de Maio de 2011, a troika, constituída por elementos da Comissão Europeia, do Banco Central Europeu (BCE) e do Fundo Monetário Internacional (FMI), apresentou o programa de assistência financeira a Portugal, que viria a vigorar até 04 de Maio de 2014. O ano de 2013 já se configurou como um ano de ténue recuperação económica, o que contribuiu para a saída em 2014 do programa de assistência financeira.

Analisando os dados anuais disponíveis no PORDATA do valor médio dos prédios transaccionados: total, urbanos e rústicos, relativos ao VMPT em Portugal por NUTSII, durante o período de 2000 a 2018, representados na Figura 3.2, observa-se que os valores médios mais elevados correspondem à NUTSII da área metropolitana de Lisboa, seguida da NUTSII do Algarve. Toda as NUTSII apresentam uma evolução do VMPT positiva, exceptuando-se a área metropolitana de Lisboa e a Região Autónoma da Madeira no ano de 2018.

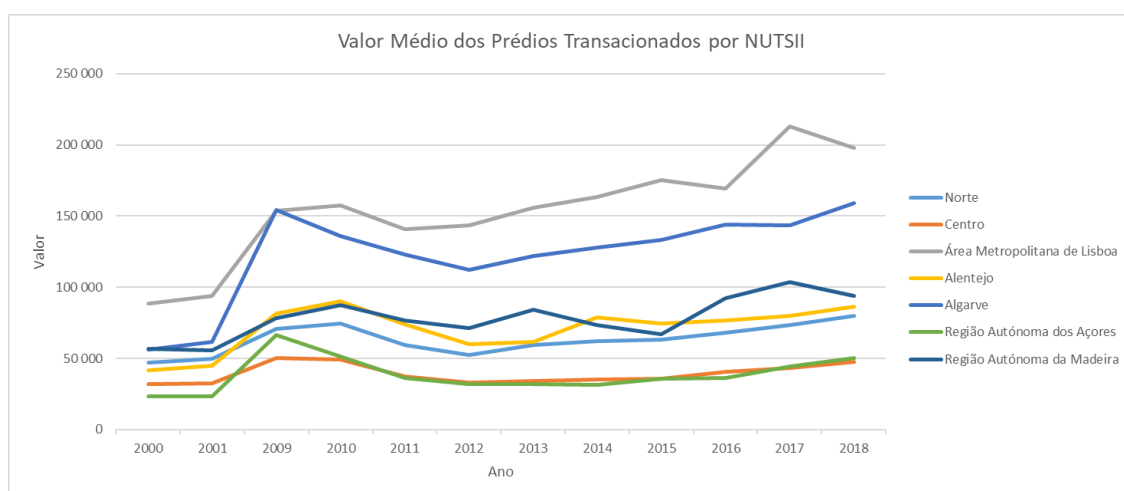


Figura 3.2: Evolução do VMPT por NUTSII

### 3.3 Dados

Os dados utilizados para o caso em estudo, foram retirados da base de dados PORDATA <sup>1</sup> e são referentes ao ano de 2017. Foram usados dados referentes a este ano em particular por serem os últimos dados disponíveis para a maioria das variáveis consideradas na literatura.

Os dados são relativos a todos os 308 municípios de Portugal, sendo que 278 são municípios no Continente, 19 são municípios na Região Autónoma dos Açores e 11 são municípios na Região Autónoma da Madeira, conforme se ilustra na Tabela 3.1.

NUTS I	Frequência	Percentagem
Continente	278	90,3
Região Autónoma da Madeira	11	3,6
Região Autónoma dos Açores	19	6,2
Total	308	100,0

Tabela 3.1: Distribuição dos Municípios por NUTS I

As variáveis relacionadas com a localização foram transformadas em variáveis ordinais, pelo que desta forma a classificação utilizada foi:

- NUTSI: 1-Continente, 2-Região Autónoma da Madeira e 3-Região Autónoma dos Açores (Tabela 3.1).
- NUTSII: 1-Alentejo, 2-Algarve, 3-Área Metropolitana de Lisboa, 4-Centro, 5-Norte, 6-Região Autónoma da Madeira e 7-Região Autónoma dos Açores (Tabela 3.2).
- NUTSIII: 1-Alentejo Centro, 2-Alentejo Litoral, 3-Algarve, 4-Alto Alentejo, 5-Alto Minho, 6-Alto Tâmega, 7-Área Metropolitana de Lisboa, 8-Área Metropolitana do Porto, 9-Ave, 10-Baixo Alentejo, 11-Beira Baixa, 12-Beiras e Serra da Estrela, 13-Cávado, 14-Douro, 15-Ilha da Madeira, 16-Ilha das Flores, 17-Ilha de Porto Santo, 18-Ilha de Santa Maria, 19-Ilha de São Jorge, 20-Ilha de São Miguel, 21-Ilha do Corvo, 22-Ilha do Faial, 23-Ilha do Pico, 24-Ilha Graciosa, 25-Ilha Terceira, 26-Lezíria

<sup>1</sup>PORDATA – Base de Dados Portugal Contemporâneo, disponível em <https://www.pordata.pt/>

do Tejo, 27-Médio Tejo, 28-Oeste, 29-Região de Aveiro, 30-Região de Coimbra, 31-Região de Leiria, 32-Tâmega e Sousa, 33-Terras de Trás os Montes e 34-Viseu, Dão e Lafões (Tabela A.1).

NUTSII	Frequência	Percentagem
Alentejo	58	18,8
Algarve	16	5,2
Área Metropolitana de Lisboa	18	5,8
Centro	100	32,5
Norte	86	27,9
Região Autónoma da Madeira	11	3,6
Região Autónoma dos Açores	19	6,2
Total	308	100,0

Tabela 3.2: Distribuição dos Municípios por NUTS II

### 3.3.1 Variável resposta

Para representar a variável resposta Valor/preço da Habitação, que neste trabalho se considera como uma característica dos municípios portugueses, considerou-se o Valor Médio dos Prédios Transaccionados (VMPT) em cada um dos municípios, conforme a Tabela 3.3.

### 3.3.2 Proxis para os fatores que influenciam o valor das habitações

Para representar os fatores identificados na revisão da literatura (Capítulo 2), as variáveis utilizadas neste trabalho serão as mais similares às descritas na revisão de literatura e que estão disponíveis no PORDATA, conforme se descreve na Tabela 3.3.

Determinantes	Descrição	Variável	Unidade de Medida
Valor/preço da habitação	Valor Médio dos Prédios Transaccionados	VMPT	€
Localização	NUTS I, II e III	NUTS I, II e III	NA
Densidade Populacional	Densidade Populacional	DP	Pessoas/ $Km^2$
Poder de Compra	Poder de Compra Per Capita	PCPC	€/pessoa
Dimensão das Famílias	Dimensão Média das Famílias	DMF	Pessoas
Desemprego	Percentagem de Desempregados Inscritos no IEFP	DIEFP	%
Qualidade Ambiental	Índice de Qualidade Ambiental	IQA	€/habitante
Envelhecimento	Índice de Envelhecimento	IE	n.º de idosos/100 jovens
Nível de Educação	Percentagem de Pessoas com Ensino Superior	PES	%
Estado Civil	Percentagem de Indivíduos Casados	PCAS	%
Habitação Própria	Percentagem de Fogos Habitados pelo Proprietário	PFHP	%
Habitações existentes	Proporção de Fogos por Habitante	PFH	n.º de fogos/habitante

Tabela 3.3: Descrição das Variáveis Consideradas

De referir que as variáveis DP, DMF, PES, PCAS e PFHP dizem respeito aos dados disponíveis dos Censos de 2011. Foram utilizados estes dados, por serem os dados disponíveis no PORDATA. Portanto considerou-se que estes continuam a representar as diferenças entre municípios até realização de novos inquéritos. Esta é também uma limitação para a não utilização de dados em painel, já que o painel ficaria incompleto no que respeita a estas variáveis.

Houve ainda necessidade de calcular algumas variáveis da Tabela 3.3 de forma a obter índices para o presente estudo, já que o PORDATA não apresentava as referidas variáveis

em proporção, mas com os valores totais, o que não refletia a realidade dos municípios, conforme apresentado na Tabela 3.4.

As variáveis são todas quantitativas contínuas, à exceção das variáveis de localização NUTS I, II e III, que são variáveis nominais.

Variável	Fórmula
IQA	$\frac{\text{Total de Gastos no Domínio de Gestão e Proteção do Ambiente}}{\text{População Residente}}$
PES	$\frac{\text{Total Pessoas com Ensino Superior}}{\text{Total Todos os Níveis de Escolaridade}}$
PCAS	$\frac{\text{Total Pessoas Casadas}}{\text{Total Todos Estados Cívís}}$
PFHP	$\frac{\text{Total Alojamentos Familiares Habitados pelo Proprietário}}{\text{Total Alojamentos Familiares}}$
PFH	$\frac{\text{Total de Tipologias dos Fogos}}{\text{População Residente}}$

Tabela 3.4: Cálculo de Variáveis

Variáveis	Obs.	Média	Desv.Padrão	Min.	Max.
VMPT	308	54.426	49.597	3.980	436.176
DP	308	299.34	828.14	5.1	7.363
PCPC	308	80.11	18.36	55.20	219.60
DMF	308	2.569	0.227	2.1	3.60
DIEFP	308	5.851	2.932	0	15.80
IQA	308	0.064	0.461	0.002	0.528
IE	308	216.93	106.73	40	733.30
PES	308	0.092	0.042	0.028	0.311
PCAS	308	0.490	0.038	0.375	0.561
PFHP	308	0.803	0.077	0.507	0.939
PFH	308	0.001	0.001	0.000	0.005

Tabela 3.5: Estatística Descritiva das Variáveis

Analisando a tabela 3.5 constatamos que a amostra contém 308 observações referentes aos municípios no ano de 2017. A média do valor das habitações transacionadas neste ano foi de aproximadamente 54426 Euros. A densidade populacional portuguesa em 2017 era em média de 299,34 pessoas/Km<sup>2</sup>. Os portugueses tinham em média um poder de compra de 80,11 €/pessoa. A dimensão média das famílias portuguesas variava entre 2,1 (min) e 3,6 (max) pessoas.

Existiam em Portugal uma média de 5,851 % de desempregados inscritos no centro de emprego, sendo que alguns municípios não tinham qualquer desempregado (min) em contraponto com outros onde a percentagem era de 15,8 %. Em Portugal a percentagem de gastos por pessoa em qualidade ambiental era em média de 0.064172 €.

A percentagem do índice de envelhecimento (n.º de idosos por cada 100 jovens) variava entre 40 % (min) e 733,3 % (max), o que demonstra a assimetria da população portuguesa, onde existem municípios maioritariamente habitados por populações jovens e outros com uma população demasiado envelhecida. Aliás a média da população portuguesa já evidencia uma população envelhecida com um valor de 216,9. A percentagem de indivíduos com ensino superior era em média de 0.09223 %, demonstrando uma população pouco instruída ao nível nacional.

Em Portugal em 2017 a percentagem de casados era em média de 49%. De salientar ainda que a percentagem média de fogos habitados pelo proprietário era de 80%, e a percentagem de fogos por habitante era em média de 0.001.

### 3.4 Fatores que influenciam o VMPT em Portugal

Tendo em conta os objetivos deste trabalho serão utilizados diversos métodos quantitativos de apoio à decisão, em particular, a Regressão Linear Múltipla e a Análise de *Clusters*. Nesta secção será apresentada uma breve descrição dos resultados obtidos pela aplicação destas metodologias.

#### 3.4.1 Modelo 1

Utilizando um modelo de regressão linear múltipla, onde o Valor Médio dos Prédios Transacionados (VMPT) é a variável resposta (conforme definido na Secção 3.3.1) e todas as restantes variáveis (conforme definido na Secção 3.3.2) são consideradas como variáveis independentes, obtém-se um modelo altamente significativo, ou seja, existe pelo menos uma variável independente com efeito significativo sobre o Valor Médio dos Prédios Transacionados. Para o tratamento e análise dos dados foi utilizado o *software SPSS*, para determinação dos fatores que influenciam os valores da habitação nos municípios em Portugal, tendo sido estimado o seguinte modelo:

$$\begin{aligned}
 VMPT = & \beta_0 + \beta_1 \times NUTSI + \beta_2 \times NUTSII + \beta_3 \times NUTSIII + \beta_4 \times DP \\
 & + \beta_5 \times PCPC + \beta_6 \times DMF + \beta_7 \times DIEFP + \beta_8 \times IQA \\
 & + \beta_9 \times IE + \beta_{10} \times PIES + \beta_{11} PCAS + \beta_{12} \times PFHP \\
 & + \beta_{13} \times PFH + u_i, i = 1, \dots, 308
 \end{aligned} \tag{3.1}$$

Na análise de resíduos deste modelo de regressão verificou-se que a nuvem de pontos dos resíduos tende a alargar com a magnitude do VMPT, pelo que a variância dos erros não é constante, pondo em causa o pressuposto de homocedasticidade. No sentido de estabilizar a variância dos erros substitui-se a variável VMPT por  $\ln(VMPT)$ . Ainda assim, como algumas das variáveis independentes eram não significativas, utilizou-se o método *stepwise*.

Model		Sum of Squares	df	Mean Square	F	Sig.
9	Regression	191,735	9	21,304	106,544	0,000
	Residual	59,587	298	0,200		
	Total	251,322	307			

Tabela 3.6: Teste ANOVA

Foi obtido um modelo de regressão significativo, ao fim de 9 passos do método *stepwise*. Este modelo apresentou um teste ANOVA com um valor de prova aproximadamente nulo, Tabela 3.6, logo é um modelo altamente significativo. Além disso observa-se que as variáveis significativas foram: PFHP, PCAS, NUTSI, IE, DIEFP, NUTSIII, NUTSII, DMF e PES, de acordo com os resultados para o modelo de regressão, apresentados na Tabela 3.7. Foram retiradas do modelo por não serem significativas as variáveis DP, PCPC, IQA e PFH.

Model		Unstandardized Coefficients		Standardized Coefficients		Collinearity Statistics		
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
9	(Constant)	14,929	0,821		18,184	0,000		
	PFHP	-3,008	0,487	-0,256	-6,180	0,000	0,464	2,156
	PCAS	-5,898	1,154	-0,245	-5,110	0,000	0,345	2,896
	NUTSI	-0,436	0,086	-0,245	-5,053	0,000	0,339	2,949
	IE	-0,001	0,000	-0,170	-2,989	0,003	0,247	4,049
	DIEFP	-0,058	0,012	-0,188	-4,931	0,000	0,549	1,821
	NUTSIII	-0,014	0,003	-0,168	-4,891	0,000	0,671	1,491
	NUTSII	-0,135	0,024	-0,255	-5,624	0,000	0,386	2,589
	DMF	0,958	0,206	0,241	4,647	0,000	0,296	3,373
PES	3,784	0,871	0,177	4,343	0,000	0,481	2,079	

Tabela 3.7: Coeficientes do modelo de regressão

Este modelo tem um  $R_a^2$  de 0,763, ou seja, as 9 variáveis independentes explicam cerca de 76,3% da variância da variável dependente  $\ln(\text{VMPT})$ .

O modelo *log-nível* ajustado pelo método *stepwise* é, então, o seguinte:

$$\begin{aligned} \ln(\text{VMPT}) = & 14,929 - 3,008 \times \text{PFHP} - 5,898 \times \text{PCAS} - 0,436 \times \text{NUTSI} \\ & - 0,001 \times \text{IE} - 0,058 \times \text{DIEFP} - 0,014 \times \text{NUTSIII} \\ & - 0,135 \times \text{NUTSII} + 0,958 \times \text{DMF} + 3,784 \times \text{PES} + u_i, i = 1, \dots, 308 \end{aligned} \quad (3.2)$$

Da Tabela 3.7 pode analisar-se a semi-elasticidade e concluir-se que, mantendo-se todas as outras variáveis constantes:

- Um aumento na variável PFHP de um ponto percentual implica uma diminuição de cerca de  $100[e^{-3.008} - 1] \approx 95,1\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável PCAS de um ponto percentual implica uma diminuição de cerca de  $99,7\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável NUTS I de um ponto (passar para a NUTS I seguinte) implica uma diminuição de cerca de  $35,3\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável IE de um idoso por 100 jovens implica uma diminuição de cerca de  $0,1\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável DIEFP de um ponto percentual implica uma diminuição de cerca de  $5,6\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável NUTS II de um ponto (passar para a NUTS II seguinte) implica uma diminuição de cerca de  $1,4\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável NUTS II de um ponto (passar para a NUTS II seguinte) implica uma diminuição de cerca de  $12,7\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável DMF de uma pessoa implica um aumento de  $160,6\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável PES de um ponto percentual implica um aumento de  $4298,7\text{€}$  no valor esperado da variável VMPT.

A análise realizada mostra que as hipóteses  $H_4$ ,  $H_5$ ,  $H_7$ ,  $H_8$ ,  $H_9$  e  $H_{10}$  ficam validadas indo de encontro à revisão de literatura efetuada. Concluí-se ainda que a localização, representada pelas NUTS I, II e III são significativas, indo de encontro à literatura de suporte à  $H_1$ .

Excetuando a localização, a variável IE ( $H_7$ ) é a variável que menos influência tem no valor das habitações (com um coeficiente standardizado de -0,170). Em sentido contrário as variáveis PFHP, PCAS e NUTS I e II, que têm uma grande influência, ainda que negativa, no valor médio de habitação dos municípios.

Só se verificou terem uma influência positiva no valor médio das habitações as variáveis DMF e PES, sendo o efeito das restantes variáveis significativas negativo.

As restantes hipóteses não foram validadas porque as variáveis correspondentes não se mostraram significativas: DP ( $H_2$ ), PCPC( $H_3$ ), IQA ( $H_6$ ) e PFH ( $H_{11}$ ).

### 3.4.2 Validação de Pressupostos e Análise de Resíduos

A estatística de *Durbin – Watson* apresenta um valor próximo de 2 (1,858), logo não há evidência de os resíduos estarem correlacionados.

Analisando os gráficos na Figura 3.3 parece que a distribuição dos resíduos se aproxima de uma distribuição normal. No histograma (à esquerda) a distribuição dos resíduos (barras azuis) parecem aproximar-se de uma distribuição normal (linha a preto). No *P-P plot* (à direita) os pontos encontram-se próximos da diagonal, logo é expectável que os resíduos apresentem uma distribuição normal. O teste K-S (Kolmogorov-Smirnov) apresenta um p-valor próximo de 0, o que implica a rejeição da normalidade, no entanto, como a amostra é grande ( $N=100 > 30$ ) pode usar-se o TLC (Teorema do Limite Central) e considerar a aproximação à distribuição normal dos resíduos.

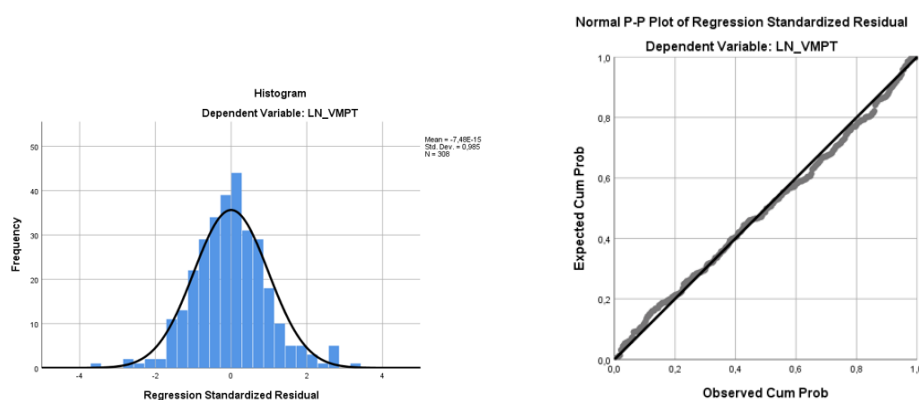


Figura 3.3: Histograma e *P – P*Plot

Os resíduos devem ser *Ruído Branco*, isto é, os resíduos devem ter média nula e variância constante. De acordo com a Figura 3.4, neste modelo a dispersão dos resíduos em torno do valor médio (zero) é, mais ou menos, aleatória. Além disso, as variâncias dos resíduos parecem ser relativamente homogêneas, ou seja, a média dos resíduos é aproximadamente nula e a variância aproximadamente constante.

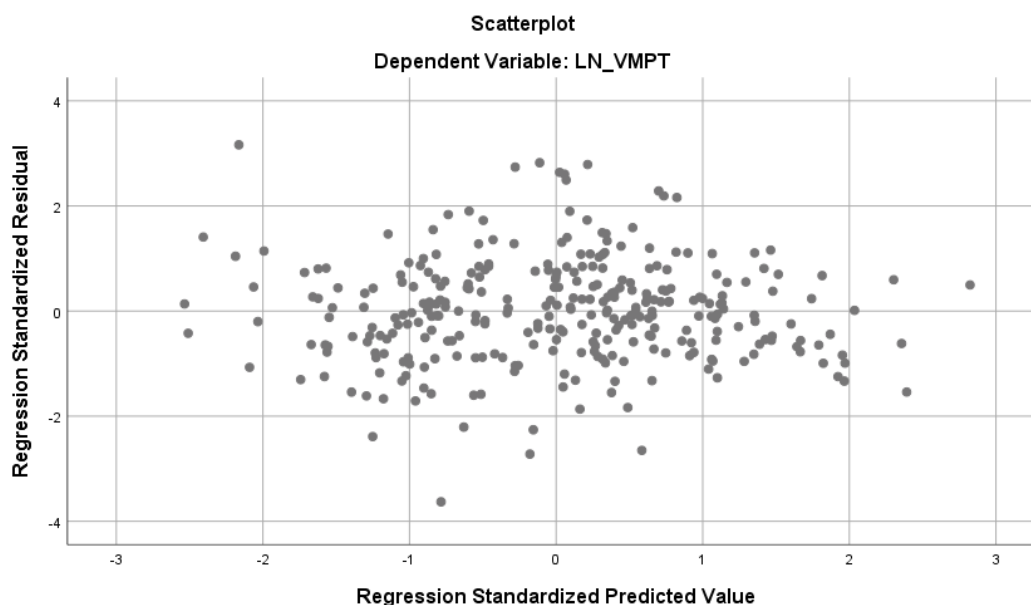


Figura 3.4: Scatterplot

O diagnóstico de colinearidade entre as variáveis pode ser feito usando os valores de T (tolerância) e os valores de VIF (*Variance Inflation Factor*), apresentados na Tabela 3.7.

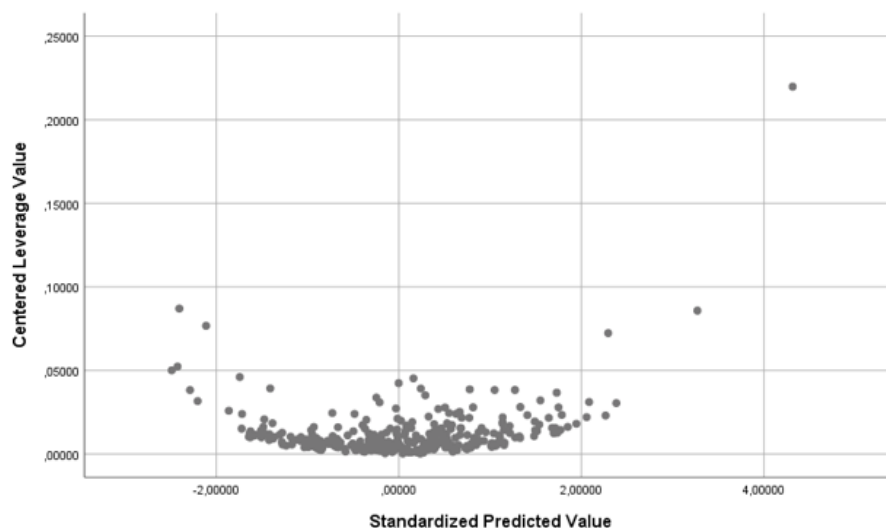
Relativamente aos valores de T, devem ser todos superiores a 0,1, o que acontece para todas as variáveis. Simultaneamente todas apresentam valores de VIF inferiores a 5, pelo que não existem evidências de existência de multicolinearidade.

Modelo	Dimension	Valores Próprios	Condition Index	Proporção de Variância										
				(Constante)	PFHP	PCAS	NUTSI	IE	DIEFP	NUTSIII	NUTSII	DMF	PES	
10*9	1	8,866	1,000	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	2	0,405	4,680	0,00	0,00	0,00	0,02	0,02	0,10	0,08	0,02	0,00	0,00	0,00
	3	0,276	5,670	0,00	0,00	0,00	0,01	0,03	0,00	0,30	0,00	0,00	0,00	0,11
	4	0,223	6,306	0,00	0,00	0,00	0,04	0,04	0,03	0,28	0,01	0,00	0,00	0,08
	5	0,119	8,642	0,00	0,00	0,00	0,00	0,10	0,31	0,02	0,08	0,00	0,16	0,16
	6	0,071	11,195	0,00	0,00	0,00	0,10	0,05	0,00	0,04	0,50	0,00	0,09	0,09
	7	0,035	15,870	0,00	0,01	0,01	0,54	0,06	0,46	0,12	0,03	0,01	0,07	0,07
	8	0,004	49,043	0,00	0,81	0,04	0,00	0,18	0,04	0,06	0,02	0,12	0,02	0,02
	9	0,002	72,954	0,00	0,02	0,75	0,05	0,47	0,02	0,07	0,01	0,50	0,00	0,00
	10	0,001	109,569	1,00	0,15	0,21	0,22	0,06	0,04	0,03	0,32	0,38	0,47	0,47

Tabela 3.8: Diagnóstico de Colinearidade

Para o Diagnóstico de Colinearidade ainda se devem analisar os valores próprios (*Eigenvalue*), os *Condition Index* e as proporções de variância, apresentados na Tabela 3.8. Os valores próprios não devem ser próximos de 0, pelo que, como os últimos valores se começam a aproximar de zero, poderá ainda haver multicolinearidade entre as variáveis consideradas. O Condition Index não deve ser superior a 30, preferencialmente inferior a 15, pelo que os valores apresentados a partir da 7.<sup>a</sup> dimensão também indicam a existência de multicolinearidade. Esse problema terá então sido resolvido pela remoção de variáveis não significativas pelo método stepwise.

No que diz respeito à existência de *outliers*, observando a Figura 3.5 verifica-se que os valores centrados de influência de cada caso no ajustamento do modelo não permitem identificar possíveis *outliers*, todas as observações encontram-se abaixo de 0,25, logo são aceitáveis.

Figura 3.5: Gráfico dos *Outliers*

Com este estudo concluiu-se portanto que, o valor médio das habitações nos municípios é afetado pela sua localização, tal com defendem Winston e Eastaway (2008) e Eddie CM Hui et al. (2007). Relativamente à dimensão média das famílias nos municípios conclui-se que esta afeta positivamente o valor das habitações, conforme definido por Varão (2019) e por Archer, Ling e Smith (2010). Confirma-se, ainda, que o desemprego afeta negativamente o valor das habitações, o que foi também apresentado por Borio e McGuire (2004) e Johnes e Hyclak (1999). O índice de envelhecimento dos municípios mostrou também ser significativo, ainda que com pouca influência no valor das habitações, tal como o sustentado por Ermisch (1996). A significância da percentagem de pessoas com ensino superior no município, de fogos habitados pelo proprietário corrobora o que apresentaram Chevan (1982) e Gyourko e Linneman (1996) e Archer, Ling e Smith 2010; Gau e Wang 1994. Assim ficou demonstrado que os fatores demográficos e o tipo de propriedade afetam o valor médio dos prédios transacionados nos municípios.

### 3.4.3 Modelo 2

Após as conclusões do primeiro modelo, procedeu-se à retirada das variáveis de localização NUTSI e NUTSIII, uma vez que estas variáveis têm um efeito intra-regional com a variável NUTS II. Ainda poderá haver relação semelhante com outras variáveis, por exemplo, Poder de Compra; Desemprego; Envelhecimento; entre outras. Esta possibilidade será averiguada aquando a análise da multicolinearidade. Assim, considera-se um segundo modelo de regressão em que apenas a variável NUTS II é considerada para representar a localização dos municípios. A variável NUTS II foi a considerada (entre as 3 representativas das NUTS), porque efetuando um teste ANOVA de comparação de médias do VMPT, por grupos definidos pelas NUTS I, II e III, apenas a variável NUTS I, não define grupos cujas médias do VMPT diferem, para um nível de significância de 5%. Assim tanto poderia ter sido considerada a NUTS II como a NUTS III, optou-se pela primeira.

Efetuiu-se novamente uma regressão linear múltipla para analisar o impacto dos fatores identificados na literatura, no valor da habitação em Portugal no ano de 2017.

Para o tratamento e análise dos dados foi utilizado o *software* R e o editor *RStudio*, para determinação dos fatores que influenciam os valores da habitação nos municípios em Portugal, tendo sido estimado o seguinte modelo:

$$\begin{aligned} \ln(\text{VMPT}) = & \beta_0 + \beta_1 \times \text{NUTSII} + \beta_2 \times \text{DP} + \beta_3 \times \text{PCPC} + \beta_4 \times \text{DMF} \\ & + \beta_5 \times \text{DIEFP} + \beta_6 \times \text{IQA} + \beta_7 \times \text{IE} + \beta_8 \times \text{PES} + \\ & \beta_9 \times \text{PCAS} + \beta_{10} \times \text{PFHP} + \beta_{11} \times \text{PFH} + u_i \end{aligned} \quad (3.3)$$

$$i = 1, \dots, 308$$

Nos modelos de regressão devem remover-se as variáveis preditoras que são altamente correlacionadas, ou seja, devem evitar-se problemas de multicolinearidade. Havendo dois fatores muito correlacionados é preferível remover um deles do modelo. Como eles fornecem informações redundantes, a remoção de um dos fatores correlacionados geralmente não reduz drasticamente o R-quadrado. Também se pode considerar a utilização do método *stepwise*, que seleciona o modelo que apresenta o maior valor de R-quadrado, ou o conhecimento especializado do conjunto de dados para remover essas variáveis.

Também se pode usar a Regressão de Mínimos Quadrados Parciais (PLS) ou a Análise de Componentes Principais que são métodos de regressão que permitem reduzir o número de variáveis preditoras a um conjunto menor de componentes não correlacionadas.

	LN(VMPT)	NUTSII	DP	PCPC	DMF	DIEFP	IQA	IE	PES	PCAS	PFHP	PFH
LN(VMPT)	1,000	-0,340	0,364	0,625	0,182	-0,076	0,018	-0,587	0,555	-0,636	-0,693	0,031
NUTSII	-0,340	1,000	0,040	-0,173	0,518	-0,306	-0,319	-0,139	-0,067	0,139	0,198	0,089
DP	0,364	0,040	1,000	0,587	0,000	0,070	-0,009	-0,216	0,564	-0,393	-0,504	-0,137
PCPC	0,625	-0,173	0,587	1,000	-0,044	-0,052	0,014	-0,404	<b>0,854</b>	-0,554	-0,671	-0,060
DMF	0,182	0,518	0,000	-0,044	1,000	-0,220	-0,312	-0,643	-0,040	-0,063	-0,116	0,103
DIEFP	-0,076	-0,306	0,070	-0,052	-0,220	1,000	0,129	0,207	-0,045	0,284	-0,082	-0,093
IQA	0,018	-0,319	-0,009	0,014	-0,312	0,129	1,000	0,167	-0,007	-0,117	-0,007	-0,001
IE	-0,587	-0,139	-0,216	-0,404	-0,643	0,207	0,167	1,000	-0,393	0,543	0,498	-0,112
PES	0,555	-0,067	0,564	<b>0,854</b>	-0,040	-0,045	-0,007	-0,393	1,000	-0,490	-0,548	-0,009
PCAS	-0,636	0,139	-0,393	-0,554	-0,063	0,284	-0,117	0,543	-0,490	1,000	0,528	-0,006
PFHP	-0,693	0,198	-0,504	-0,671	-0,116	-0,082	-0,007	0,498	-0,548	0,528	1,000	0,113
PFH	0,031	0,089	-0,137	-0,060	0,103	-0,093	-0,001	-0,112	-0,009	-0,006	0,113	1,000

Tabela 3.9: Matriz de Correlações

Calculando as correlações (Tabela 3.9) entre todas as variáveis consideradas verificou-se que as variáveis PES e PCPC estavam altamente correlacionadas, com uma correlação de cerca de 86%, existindo assim multicolinearidade entre elas.

Desta forma optou-se pela retirada da variável PES, passando o modelo em análise a ser o seguinte:

$$\begin{aligned} \ln(\text{VMPT}) = & \beta_0 + \beta_1 \times \text{NUTSII} + \beta_2 \times \text{DP} + \beta_3 \times \text{PCPC} + \beta_4 \times \text{DMF} \\ & + \beta_5 \times \text{DIEFP} + \beta_6 \times \text{IQA} + \beta_7 \times \text{IE} + \\ & \beta_8 \times \text{PCAS} + \beta_9 \times \text{PFHP} + \beta_{10} \times \text{PFH} + u_i \end{aligned} \quad (3.4)$$

$$i = 1, \dots, 308$$

Após a retirada da variável correlacionada iniciou-se o modelo de estimação do OLS apresentado na Tabela 3.10. Constatou-se que o valor de prova do teste global é  $p\text{-value} < 2.2e-16$ , então para um nível de significância de 5%, rejeita-se a hipótese nula, logo existe pelo menos um coeficiente de regressão que é significativamente diferente de zero, ou seja, existe pelo menos uma variável independente que contribui para explicar a variável dependente. Da análise aos testes marginais, verificou-se que quase todas as variáveis são significativas para um nível de significância de 10%, à exceção de DP e IQA. Além disso o modelo apresenta um  $R^2$  de 0,7121, ou seja, as variáveis independentes explicam cerca de 71,21% da variância da variável dependente VMPT.

Coefficients	Estimate	Std. Error	t value	Pr(> t )	Sig. <sup>1</sup>
(Intercept)	1,35E+01	8,94E-01	15,138	< 0,0001	***
NUTSII	-2,13E-01	2,20E-02	-9,67	< 0,0001	***
DP	2,59E-05	4,52E-05	0,572	0,5674	
PCPC	9,83E-03	2,42E-03	4,07	< 0,0001	***
DMF	1,00E+00	2,17E-01	4,616	< 0,0001	***
DIEFP	-1,87E-02	1,13E-02	-1,655	0,0989	.
IQA	-7,30E-01	6,58E-01	-1,111	0,2686	
IE	-9,06E-04	5,03E-04	-1,803	0,0724	.
PCAS	-5,44E+00	1,15E+00	-4,746	< 0,0001	***
PFHP	-3,27E+00	5,62E-01	-5,825	< 0,0001	***
PFH	8,86E+01	4,09E+01	2,167	0,031	*

<sup>1</sup> Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 30800 on 297 degrees of freedom

Multiple R-squared: 0.7215, Adjusted R-squared: 0.7121

F-statistic: 76.94 on 10 and 297 DF, p-value: < 2.2e-16

Tabela 3.10: Modelo de Estimação OLS

Passou-se então à validação dos pressupostos da normalidade dos erros, ausência de autocorrelação dos erros e homocedasticidade, para verificar se a inferência estatística é válida. Se os resíduos  $u_i$  são normalmente distribuídos, os estimadores OLS também serão normalmente distribuídos.

#### 3.4.4 Validação de Pressupostos e Análise de Resíduos

Para verificação do pressuposto da homocedasticidade, realizou-se o teste de *Breusch – Pagan* para o qual se obteve uma estatística de teste de 45,694 e um valor de prova de 0,1269. Uma vez que o valor de prova não é inferior a 0,05, não se rejeita a hipótese nula. Assim, não há evidências estatísticas para dizer que a heterocedasticidade está presente no modelo de regressão.

Para a validação do pressuposto da normalidade dos erros, realizou-se o teste *Shapiro – Wilk*. O valor de prova encontrado (0.002582) é um valor inferior a 0,01, logo rejeita-se a hipótese nula com um grau de significância de 1%, ou seja, por este teste poder-se-ia concluir que os erros não são normalmente distribuídos.

No entanto como a dimensão da amostra é superior a 100, pode assumir-se uma aproximação à distribuição normal. Além disso, pode-se observar essa aproximação pelo análise ao histograma e ao *QQ-Plot* (Figura 3.6). Observando o *QQ-Plot* não parece haver um desvio grande nos resíduos relativamente à linha e ao intervalo de confiança admissível. Parece que

os resíduos têm distribuição normal. O mesmo se verifica no histograma dos resíduos cujas barras sugerem uma distribuição que tem semelhanças com a linha da distribuição normal, Figura 3.6.

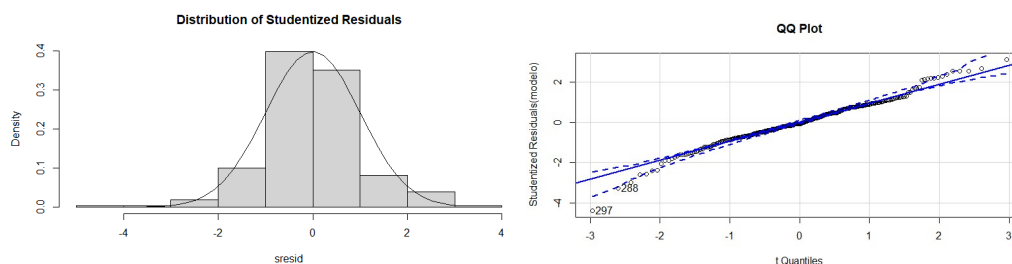


Figura 3.6: Histograma e P-P Plot

Para verificação do pressuposto da ausência de autocorrelação efetuou-se o teste de *Durbin–Watson*. O resultado obtido apresenta um  $p$ -value  $< 0,002$  o que pressupõe a rejeição da hipótese nula, para um nível de significância de 1%, donde podemos concluir que os resíduos não são independentes ou estão correlacionados.

Para verificar se o modelo inclui variáveis explicativas linearmente independentes (ausência de multicolinearidade), calcularam-se os valores do VIF para cada uma das variáveis independentes no modelo. O modelo pode apresentar problemas em duas das suas variáveis independentes (DMF=3,173766 e IE=3,749748) para as quais os valores de VIF são superiores a 3.

De seguida verificou-se a existência de *outliers*, usando o teste de *Bonferroni*, cuja hipótese nula é  $H_0$ : Não existem *outliers* severos. O  $p$ -value obtido é de  $1.7075e-05 < 1\%$ , pelo que não existem evidências da presença de outliers, para este nível de significância.

Como forma de resolver os problemas observados nos pressupostos, procedeu-se à análise do modelo de estimação GLS (*Generalized Least Squares*), uma vez que se trata do modelo que por norma valida os pressupostos da autocorrelação e homocedasticidade dos resíduos.

Os resultados obtidos são os constantes da Tabela 3.11.

O modelo estimado através do GLS pode ser representado da seguinte forma:

$$\begin{aligned} \ln(\text{VMPT}) = & 10.7411 - 0.1269 \times \text{NUTSII} + 0.00003 \times \text{DP} + 0.0045 \times \text{PCPC} \\ & + 0.3344 \times \text{DMF} - 0.0157 \times \text{DIEFP} + 0.3530 \times \text{IQA} - 0.0002 \times \text{IE} \quad (3.5) \\ & + 0.4172 \times \text{PCAS} - 1.5682 \times \text{PFHP} + 62.0453 \times \text{PFH} + u_i, i = 1, \dots, 308 \end{aligned}$$

Fazendo a análise da Tabela pode concluir-se que, mantendo-se todas as outras variáveis constantes:

- Um aumento na variável NUTS I de um ponto (passar para a NUTS I seguinte) implica uma diminuição de cerca de  $100[e^{-0.12690} - 1] \approx 11,92\text{€}$  no valor esperado da variável VMPT;
- Um aumento na variável PCPC de um euro por pessoa implica um aumento de cerca de  $0,45\text{€}$  no valor esperado da variável VMPT;

Coefficients	Estimate	Std. Error	t value	Pr(> t )	Sig. <sup>1</sup>
(Intercept)	10.74110	0.709019	15.149248	0.0000	***
NUTSII	-0.12690	0.032897	-3.857496	0.0001	***
DP	0.00003	0.000026	1.136936	0.2565	
PCPC	0.00454	0.001434	3.163410	0.0017	**
DMF	0.33442	0.150389	2.223719	0.0269	*
DIEFP	-0.01575	0.008228	-1.914028	0.0566	
IQA	0.35309	0.341105	1.035123	0.3015	
IE	-0.00021	0.000298	-0.691466	0.4898	
PCAS	0.41720	0.856603	0.487036	0.6266	
PFHP	-1.56822	0.353665	-4.434196	0.0000	***
PFH	62.04535	22.197985	2.795089	0.0055	**

<sup>1</sup> Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

Residual standard error: 0.3441522

Degrees of freedom: 308 total; 297 residual

Tabela 3.11: Modelo de Estimação GLS

- Um aumento na variável DMF de uma pessoa implica um aumento de cerca de 39,71€ no valor esperado da variável VMPT;
- Um aumento na variável PFHP de um ponto percentual implica uma diminuição 79,16€ no valor esperado da variável VMPT.
- Um aumento na variável PFH de um fogo por habitante implica um aumento de cerca de 8,83€ no valor esperado da variável VMPT;

para as restantes variáveis não é possível fazer uma interpretação, pois são não significativas.

A análise realizada mostra que as hipóteses  $H_3$ ,  $H_4$ ,  $H_{10}$  e  $H_{11}$  ficam validadas indo de encontro à revisão de literatura efetuada. Conclui-se ainda que a localização, representada pela NUTS II é significativa, indo de encontro à literatura de suporte à  $H_1$ , demonstrando que a localização apesar dos possíveis efeitos intra-regionais se mantêm como significativa.

A variável PCPC ( $H_3$ ) é a variável que menos influência tem no valor das habitação. Em sentido contrário a variável PFH, tem a maior influência, no valor médio das habitações dos municípios.

As variáveis PCPC ( $H_3$ ), DMF ( $H_4$ ) e PFH ( $H_{11}$ ) verificou-se terem uma influência positiva no valor médio das habitações, em sentido contrário as restantes variáveis significativas NUTSII ( $H_1$ ) e PFHF ( $H_{10}$ , apresentam um efeito negativo.

As restantes hipóteses não foram validadas porque as variáveis correspondentes não se mostraram significativas: DP ( $H_2$ ), DIEFP ( $H_5$ ), IQA ( $H_6$ ), IE ( $H_7$ ), PES ( $H_8$ ) e PCAS ( $H_9$ ).

Conclui-se portanto, após a segunda análise efetuada que, o valor médio das habitações nos municípios é afetado pela sua localização, tal com defendem Winston e Eastaway (2008), Kaya e Atan (2014) e Alkana et al. (2015). Relativamente ao poder de compra per capita conclui-se que este afeta positivamente o valor das habitações, conforme definido por Calcagno, Fornero e Rossi (2009), Fuller, Johnston e Regan (2020) e S. McGreal e Taltavull de La Paz (2013). A dimensão média das famílias nos municípios afeta positivamente o valor das habitações, conforme definido por Archer, Ling e Smith (2010) e Varão (2019).

Confirma-se, ainda, que a percentagem de fogos habitados pelo proprietário afeta negativamente o valor das habitações, o que foi também apresentado por Archer, Ling e Smith (2010) e Gau e Wang (1994). A significância da percentagem de fogos por habitante no município, corrobora Katz e Rosen (1987) e Xu (2017), demonstrando assim que o tipo de propriedade afeta o valor médio dos prédios transacionados nos municípios.

Assim, os modelos de regressão estimados (Modelo 1 e 2) são válidos, podendo apresentar-se resumidamente os resultados na Tabela 3.12. Fazendo uma comparação entre as duas análises podemos concluir que as variáveis que se mantêm significativas são a localização  $H_1$ , o que também conseguimos confirmar através da análise à figura 3.2, onde conseguimos concluir que os valores são significativamente diferentes dependendo da região, a dimensão das famílias  $H_4$ , o que faz crer que as famílias adquirem a sua habitação de acordo com as suas necessidades de quartos, e o tipo de posse da propriedade  $H_{10}$  que afeta negativamente o valor, demonstrando que é pouco provável as famílias adquirirem uma segunda habitação, quando já são proprietários.

Hipótese Formulada	Modelo 1	Modelo 2
H1 - A localização afeta o valor das habitações	Validada	Validada
H2 - A densidade populacional afeta o valor das habitações	Não Validada	Não Validada
H3 - O poder de compra afeta o valor das habitações	Não Validada	Validada
H4 - A dimensão das famílias afeta o valor das habitações	Validada	Validada
H5 - O desemprego afeta o valor das habitações	Validada	Não Validada
H6 - A qualidade ambiental afeta o valor das habitações	Não Validada	Não Validada
H7 - O envelhecimento afeta o valor das habitações	Validada	Não Validada
H8 - O nível de educação afeta o valor das habitações	Validada	Não Validada
H9 - Os fatores demográficos afetam o valor das habitações	Validada	Não Validada
H10 - O tipo de propriedade afeta o valor das habitações	Validada	Validada
H11 - O número de habitações afeta o valor das habitações	Não Validada	Validada

Tabela 3.12: Confirmação das Hipóteses

Assim, apenas a densidade populacional e a qualidade ambiental dos municípios não se mostraram significativas para o valor médio das habitações transacionadas, em ambos os modelos. No entanto, o modelo 2 ainda pode ser afinado, retirando variáveis não significativas, por exemplo a variável PCAS, que possui o menor valor da estatística de teste  $t$ , o que pode implicar a validação de outras hipóteses.

### 3.5 Identificação de Clusters de Municípios Homogêneos

Neste trabalho, pretende também estudar-se se existem semelhanças entre municípios, no que respeita aos fatores que influenciam os valores das habitações em Portugal e averiguar se essas diferenças/semelhanças se manifestam efetivamente no VMPT (O3). Para tal efetuou-se uma análise de *clusters* no sentido de detetar grupos homogêneos nos dados, conforme se descreveu sucintamente na Secção 2.4.2.

Utilizando o *software* IBM SPSS, versão 25, efetuou-se uma análise combinada entre um método hierárquico e um método não hierárquico de *clusters*.

Na análise hierárquica utilizou-se a distância euclidiana, com o método de agrupamento distância entre grupos (*Between-groups linkage*), tendo-se obtido o dendrograma na Figura B.1, do Apêndice B. As variáveis consideradas nesta análise hierárquica foram as mesmas que foram utilizados no segundo modelo da regressão linear, excluem-se portanto as NUTSI e III. Foi ainda considerada a transformação das variáveis, escolhendo a opção *Standardize*:

Iteration History <sup>a</sup>		
Iteration	Centers	
	1	2
1	5,182	5,635
2	0,031	0,806
3	0,035	0,922
4	0,038	0,728
5	0,063	0,857
6	0,062	0,632
7	0,058	0,510
8	0,055	0,410
9	0,031	0,217
10	0,057	0,344

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is ,202. The current iteration is 10. The minimum distance between initial centers is 15,759.

Iteration History <sup>a</sup>			
Iteration	Change in Cluster Centers		
	1	2	3
1	5,370	4,570	5,620
2	0,060	1,578	1,414
3	0,042	0,417	0,873
4	0,047	0,527	0,634
5	0,063	0,700	0,679
6	0,190	0,838	1,043
7	0,232	0,738	0,585
8	0,264	0,509	0,488
9	0,169	0,562	0,248
10	0,121	0,141	0,154

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is ,079. The current iteration is 10. The minimum distance between initial centers is 12,441.

Iteration History <sup>a</sup>				
Iteration	Change in Cluster Centers			
	1	2	3	4
1	5,237	4,860	4,665	4,885
2	0,984	0,230	0,263	0,000
3	0,417	0,048	0,052	0,000
4	0,000	0,012	0,013	0,000
5	0,000	0,023	0,025	0,000
6	0,000	0,012	0,013	0,000
7	0,000	0,000	0,000	0,000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 7. The minimum distance between initial centers is 9,936.

Figura 3.7: Convergência

Z scores. No dendrograma (Figura B.1, do Apêndice B) pode observar-se que há muitos municípios homogêneos, havendo depois um grupo mais pequeno de municípios que se diferencia muito dos restantes. Desta Análise de *Clusters* hierárquica pode, então, destacar-se a formação de dois *clusters*, no entanto não se afasta a possibilidade da existência de mais *clusters*.

Assim na análise não hierárquica (*k-means*) consideram-se 2, 3 e 4 *clusters*, para depois averiguar qual o número de *clusters* mais adequado. As variáveis consideradas nesta análise foram as mesmas que na estratégia anterior.

Inicia-se o processo de aplicação da metodologia por um processo de normalização ou standardização das variáveis. Correndo o algoritmo com um número máximo de 10 iterações e com o critério de convergência 0, para um número de *clusters*  $k = 2, 3$  e 4, obtem-se convergência apenas para  $k = 4$ , por já não ocorrerem variações nos centróides dos *clusters*, como se pode observar na Figura 3.7. Assim, no que respeita à convergência seria mais adequado considerar 4 *clusters*.

## ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(DP)	89,376	1	,711	306	125,671	,000
Zscore(PCPC)	134,337	1	,564	306	238,078	,000
Zscore(DMF)	1,487	1	,998	306	1,490	,223
Zscore(DIEFP)	1,977	1	,997	306	1,984	,160
Zscore(IQA)	,035	1	1,003	306	,035	,852
Zscore(IE)	28,913	1	,909	306	31,815	,000
Zscore(PES)	147,061	1	,523	306	281,362	,000
Zscore(PCAS)	93,469	1	,698	306	133,946	,000
Zscore(PFHP)	92,455	1	,701	306	131,867	,000
Zscore(PFH)	,449	1	1,002	306	,448	,504
Zscore(NUTSII)	5,389	1	,986	306	5,468	,020

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Figura 3.8: Testes ANOVA para  $k = 2$  clusters

As distâncias entre os centroides finais considerando 2 clusters é de 3,865. Considerando 3 clusters tem-se que as distâncias são: 5,134; 4,005 e 2,644. Com 4 clusters as distâncias vão de 2,789 a 9,579.

Relativamente aos testes ANOVA, na Figura 3.8 pode observar-se que no caso de  $k = 2$ , a maioria das variáveis (Zscore(DMF), Zscore(DIEFP), Zscore(IQA), Zscore(PFH), Zscore(NUTSII)) não têm média significativamente diferente por cluster.

Na Figuras D.1 e D.2, em anexo, os testes ANOVA indicam que para  $k = 3$  e  $k = 4$  as variáveis são todas significativas, ou seja, as suas médias diferem significativamente por cluster. Assim, poderá ser mais adequado considerar  $k = 3$  ou  $k = 4$  para estas variáveis.

Em virtude da análise de clusters ser uma técnica multivariada que não possui sólidos fundamentos teóricos, uma vez que apenas agrupa casos segundo critérios mais ou menos heurísticos, carece de validação através de outras análises complementares, por exemplo a análise discriminante. Também podem ser consideradas medidas de validação, tais como as apresentadas na Secção 2.4.2.

De acordo com o descrito na Secção 2.4.2 os clusters resultantes devem ter boas propriedades estatísticas, como:

- testes ANOVA significativos – o que neste caso aconteceu para  $k = 3$  e  $k = 4$  (Figura 3.8),
- Convergência do algoritmo – o que neste caso aconteceu para  $k = 4$  (Figura 3.7),
- clusters com número homogêneo de casos – o que neste caso não aconteceu em nenhum caso, a distribuição mais equilibrada é para  $k = 3$  (Figura 3.11),
- ter medidas de validação o mais otimizadas possível.

O cálculo das medidas de validação não está disponível no SPSS, pelo que a validação dos resultados foi feita usando o package *cValid* do software *R*. Considerou-se o método *kmeans* para 2, 3 e 4 clusters e as medidas de validação internas e de estabilidade. Foi usada a distância euclídeana e o método *average*. As medidas de validação obtidas são as apresentadas na Tabela 3.13. As medidas ótimas são as apresentadas a negrito na Tabela 3.13 e sumariamente na Tabela 3.14. As medidas de estabilidade devem ser todas minimizadas, bem como a conectividade (*Connectivity*) e a outras medidas de validação (*Dunn* e *Silhouette*) maximizadas.

Método	Medida	Número de Clusters		
		2	3	4
kmeans	APN	<b>0.1178</b>	0.2032	0.2181
	AD	3.9453	3.7438	<b>3.5497</b>
	ADM	<b>0.4716</b>	0.8469	0.9144
	FOM	0.9408	0.9085	<b>0.8876</b>
	<i>Connectivity</i>	<b>77.2960</b>	90.4611	95.1190
	<i>Dunn</i>	0.0483	<b>0.0768</b>	0.0626
	<i>Silhouette</i>	<b>0.2263</b>	0.1838	0.1991

Tabela 3.13: Medidas de Validação dos Clusters

Medida	Score	Método	Número de <i>Clusters</i>
APN	0.1178	kmeans	2
AD	3.5497	kmeans	4
ADM	0.4716	kmeans	2
FOM	0.8876	kmeans	4
Connectivity	77.2960	kmeans	2
Dunn	0.0768	kmeans	3
Silhouette	0.2263	kmeans	2

Tabela 3.14: Optimal Scores

Assim, genericamente e tendo em conta as medidas ótimas, parece ser mais adequado considerar 2 clusters, Tabelas 3.13 e 3.14. Na Figura 3.9 ainda se pode ver a representação gráfica das medidas de validação interna e de estabilidade.

Variáveis	Cluster	
	1	2
Zscore(DP)	-0,229	1,269
Zscore(PCPC)	-0,280	1,556
Zscore(DMF)	0,029	-0,164
Zscore(DIEFP)	-0,034	0,189
Zscore(IQA)	0,005	-0,025
Zscore(IE)	0,130	-0,722
Zscore(PES)	-0,293	1,628
Zscore(PCAS)	0,234	-1,298
Zscore(PFHP)	0,232	-1,291
Zscore(PFH)	0,016	-0,090
Zscore(NUTSII)	0,056	-0,312

	Cluster		
	1	2	3
	-0,110	1,678	-0,322
	0,033	1,900	-0,547
	0,623	-0,304	-0,640
	-0,379	0,196	0,386
	-0,318	-0,001	0,368
	-0,550	-0,690	0,821
	0,033	1,938	-0,557
	-0,137	-1,396	0,532
	-0,074	-1,482	0,483
	0,234	-0,179	-0,223
	0,451	-0,370	-0,423

	Cluster			
	1	2	3	4
	4,861	0,006	-0,316	-0,291
	2,979	0,356	-0,587	-0,373
	-0,597	0,491	-0,513	-0,158
	0,328	-0,407	0,400	1,779
	0,497	-0,267	0,138	6,092
	-0,577	-0,626	0,755	-0,487
	2,858	0,338	-0,556	-0,489
	-1,831	-0,422	0,611	-0,719
	-2,147	-0,362	0,543	0,192
	-0,765	0,180	-0,140	-0,626
	-0,083	0,262	-0,255	-1,646

Tabela 3.15: Centróides Finais dos Clusters

Os Centróides Finais dos *Clusters* são apresentados na Tabela 3.15 e ilustrados na Figura 3.10 para  $k = 2$ . Para  $k = 3$  e  $k = 4$  podem ser vistos no Apêndice Figura C.1 e Figura C.2. Observando a Tabela 3.15 e o gráfico da Figura 3.10 pode observar-se que, quando se consideram 2 *clusters* o *cluster* 1 é composto pelos municípios com densidade populacional abaixo da média, enquanto que o *cluster* 2 é composto pelos municípios com maior densidade populacional. Este segundo *cluster* é constituído nomeadamente por os municípios da área metropolitana do Porto e Lisboa e a maioria das capitais de distrito, conforme se apresenta no Apêndice E. O *cluster* 1 é composto por 261 municípios, enquanto que o *cluster* 2 é composto por 47 municípios, demonstrando uma grande disparidade entre eles, Figura 3.11. A distribuição por três *clusters* é bastante mais equilibrada sugerindo ser mais adequado considerar este número de *clusters*. Para 4 *clusters* a divisão também é bastante dispare.

No *cluster* 1 estão os municípios onde a média da variável localização (NUTS II), dimensão média das famílias (DMF), índice de qualidade ambiental (IQA), índice de envelhecimento (IE), percentagem de casados (PCAS), percentagem de fogos habitados pelo proprietário (PFHP) e percentagem de fogos por habitante (PFH) estão acima da média global, e a densidade populacional (DP), poder de compra per capita (PCPC), percentagem de desempregados inscritos no instituto de emprego e formação profissional (DIEFP), assim como a percentagem de habitantes com o ensino superior (PES) estão abaixo da média global, o que parece corresponder aos municípios mais pobres e menos populosos.

O *clusters* 2 é composto pelos municípios onde a média da variável densidade populacional (DP), poder de compra per capita (PCPC), desempregados inscritos no instituto do emprego

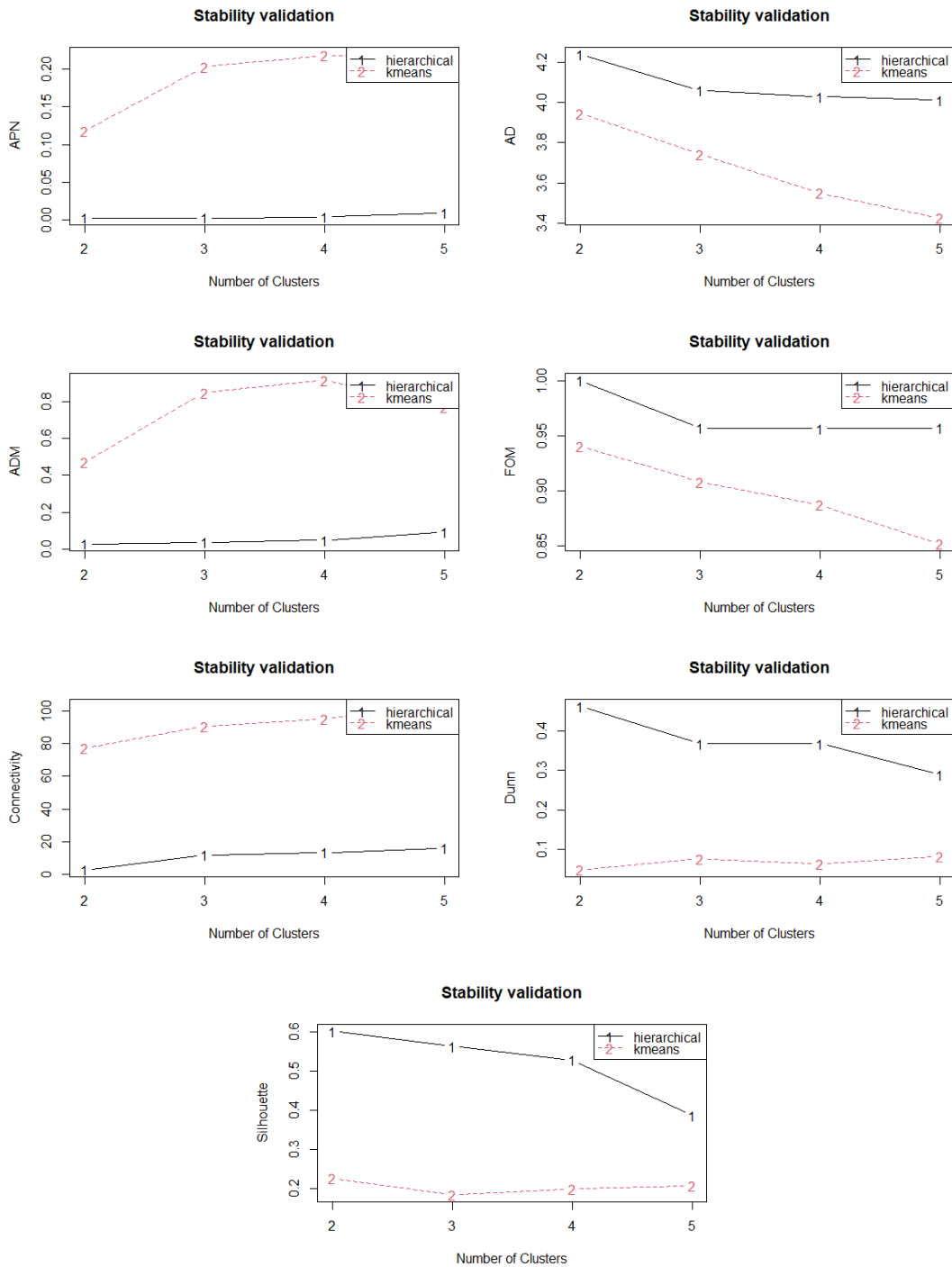


Figura 3.9: Representação das medidas de validação

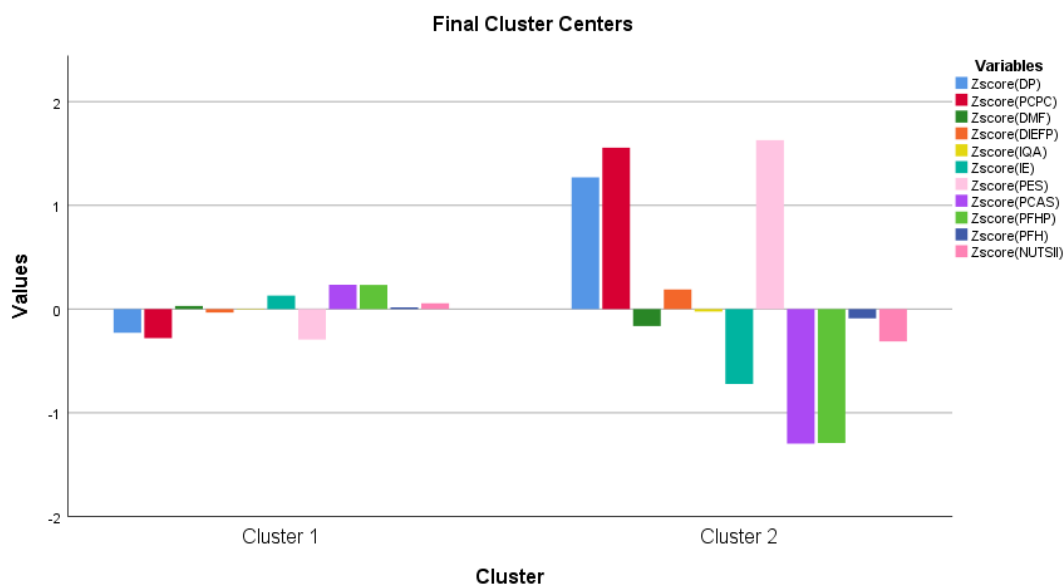


Figura 3.10: Representação dos centróides finais para  $k = 2$

Number of Cases in each Cluster		Number of Cases in each Cluster		Number of Cases in each Cluster	
Cluster	1	261,000	Cluster	1	147,000
	2	47,000		2	34,000
Valid		308,000		3	127,000
Missing		0,000	Valid		308,000
			Missing		0,000
			Cluster	1	9,000
				2	157,000
				3	139,000
				4	3,000
			Valid		308,000
			Missing		0,000

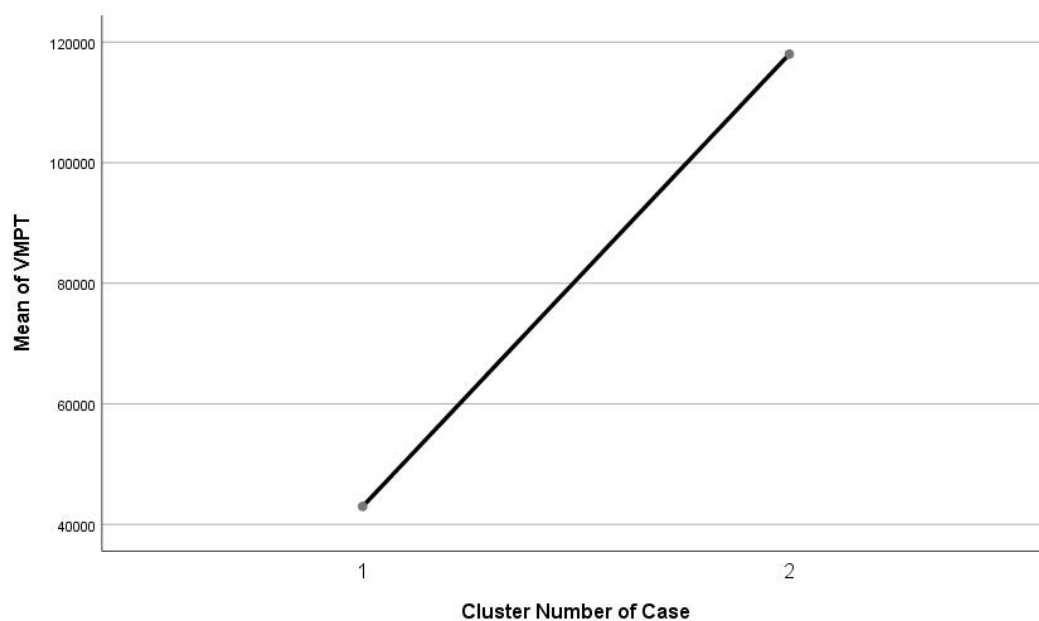
Figura 3.11: Número de Municípios por *cluster*

e formação profissional, percentagem de habitantes com o ensino superior estão acima da média global, e a localização (NUTSII), dimensão média das famílias (DMF), Índice de qualidade ambiental (IQA), índice de envelhecimento (IE), percentagem de casados (PCAS), percentagem de fogos habitados pelo proprietário (PFHP) e percentagem de fogos por habitante (PFH) estão abaixo da média global, o que parece corresponder aos municípios mais ricos e mais populosos.

No que diz respeito ao valor médio dos prédios transacionados observam-se na Figura 3.12 e as dos anexos constantes do apêndice C, diferenças de médias entre os municípios, seja qual for o número de *clusters* considerado. Os preços são significativamente diferentes para os *clusters* em todos os casos, de acordo com os resultados de testes ANOVA para a comparação destes valores médios por *cluster*.

Os gráficos para as médias do VMPT para 3 e 4 *clusters* podem ser observados no Apêndice C e os municípios que os constituem nos Apêndices F e G. E uma análise semelhante à que foi feita para  $k = 2$  *clusters*, pode ser feita para  $k = 3$  e  $k = 4$ .

Para  $k = 3$  *clusters* (Apêndice F e Figura C.1) o *cluster* 2 é o que se diferencia mais dos outros 2, com densidade populacional (DP), poder de compra per capita (PCPC) e percentagem de pessoas com o ensino superior (PES) muito acima da média global e índice

Figura 3.12: Médias dos VMPT por *cluster*

de envelhecimento (IE), percentagem de indivíduos casados (PCAS) e a percentagem de fogos habitados pelo proprietário (PFHP) abaixo da média. Neste *cluster* o VMHT também é superior ao dos outros 2 *clusters* e inclui os municípios mais industrializados do país.

Para  $k = 4$  *clusters* (Apêndice G e Figura C.2) o *cluster* 1 é o que se diferencia mais dos outros 3, com densidade populacional (DP), poder de compra per capita (PCPC) e percentagem de pessoas com o ensino superior (PES) muito acima da média global e percentagem de indivíduos casados (PCAS) e a percentagem de fogos habitados pelo proprietário (PFHP) abaixo da média. Neste *cluster* o VMHT também é superior ao dos outros 3 *clusters* e inclui os municípios de Matosinhos, Porto, São João da Madeira, Almada, Amadora, Cascais, Lisboa, Odivelas, Oeiras, municípios bastante citadinos.

O *cluster* 4 diferencia-se por ter um índice de qualidade ambiental (IQA) e uma percentagem de desempregados inscritos no IEFP (IQA) muito superiores à média. O VMHT é o mais baixo dos 4 *clusters* e inclui os municípios de Barrancos, Cartaxo e Mourão, municípios de características rurais.

Em conclusão e tal como foi referido na Secção 2.4.2 não é possível dizer de forma definitiva o número de *clusters* mais adequados. No entanto, como o teste ANOVA para 2 *clusters* (Figura 3.8) revelou que nem todas as variáveis são significativas, se o decisor quiser ter em linha de conta todos os determinantes referidos na literatura deve considerar a divisão em 3 *clusters*, tanto na definição de políticas de apoio à habitação como na decisão de adquirir uma habitação. Se não for esse o caso as medidas de validação sugerem a adopção de 2 *clusters* (Tabelas 3.13 e 3.14) com as características já descritas. Em qualquer dos casos os municípios mais metropolitanos destacam-se com VMPT mais elevado.



## Capítulo 4

# Conclusões, Contribuições, Limitações e Trabalho Futuro

O mercado imobiliário português tem vivido uma época de expansão, com os valores das casas a registarem subidas significativas. No entanto, os primeiros sinais de abrandamento começam a fazer-se sentir, com uma subida bastante mais gradual. Reflexo do abrandamento da procura, da falta de oferta suficiente, ou da incapacidade do mercado português para conseguir pagar os valores praticados, a realidade é que, embora os valores das casas continuem a ser elevados, o ritmo de subida está mais lento o que leva a entender que o mercado está a ajustar-se à realidade portuguesa. Este estudo pretende dar um contributo para a tomada de decisão ao nível individual, nomeadamente na ajuda da tomada de decisão quer dos compradores, quer dos investidores no mercado imobiliário residencial, e ao nível dos decisores políticos municipais, intermunicipais ou nacionais no apoio à tomada de decisão das políticas públicas de habitação. O estudo utilizou a regressão linear múltipla, com o objectivo de analisar o impacto dos fatores identificados na literatura no valor das habitações dos municípios em Portugal no ano de 2017, tendo também efetuado uma análise de *clusters* para verificar a existência de semelhanças entre municípios, no que respeita aos factores que influenciam o valor das habitações em Portugal e averiguar se essas diferenças/semelhanças se manifestam efetivamente no VMPT. Os dados utilizados, foram retirados da base de dados PORDATA.

Numa primeira análise de regressão linear (modelo 1) concluiu-se que, o valor médio das habitações nos municípios é afetado pela sua localização, tal com defendem Winston e Eastaway (2008). Relativamente à dimensão média das famílias nos municípios conclui-se que esta afeta positivamente o valor das habitações, conforme definido por Varão (2019). Confirma-se, ainda, que o desemprego afeta negativamente o valor das habitações, o que foi também apresentado por Borio e McGuire (2004). O índice de envelhecimento dos municípios mostrou também ser significativo, ainda que com pouca influência no valor das habitações, tal como o sustentado por Ermisch (1996). A significância da percentagem de pessoas com ensino superior no município, de fogos habitados pelo proprietário corrobora Archer, Ling e Smith (2010), Chevan (1982), Gau e Wang (1994) e Gyourko e Linneman (1996), demonstrando assim que os fatores demográficos e o tipo de propriedade afetam o valor médio dos prédios transacionados nos municípios.

Verificado o primeiro modelo, procedeu-se à retirada das variáveis de localização NUTSI e NUTSIII, uma vez que estas variáveis têm um efeito intra-regional com a variável NUTS II. Assim, considerou-se um segundo modelo de regressão em que apenas a variável NUTS II foi considerada para representar a localização dos municípios.

Neste modelo saiu reforçada a conclusão segundo a qual, o valor médio das habitações nos municípios é afetado pela sua localização, tal com defendem Alkana et al. (2015), Kaya e Atan (2014) e Winston e Eastaway (2008). Concluiu-se que o valor das habitações é afetado positivamente pelo poder de compra per capita, tal como o investigada por Calcagno, Fornero e Rossi (2009), Fuller, Johnston e Regan (2020) e S. McGreal e Taltavull de La Paz (2013), e pela dimensão média das famílias nos municípios, de acordo com as conclusões de Archer, Ling e Smith (2010) e Varão (2019). Foi ainda possível concluir, que a percentagem de fogos habitados pelo proprietário afeta negativamente o valor das habitações, o que foi também apresentado por Gau e Wang (1994), Archer, Ling e Smith (2010). A significância da percentagem de fogos por habitante no município, corrobora Xu (2017), Katz e Rosen (1987), demonstrando assim o tipo de propriedade afetam o valor médio dos prédios transacionados nos municípios.

Na análise comparativa dos dois modelos, constatamos que as variáveis que se mantêm significativas a ambos, são a localização  $H_1$ , o que nos permite concluir que os valores são significativamente diferentes, dependendo da região do país. As zonas metropolitanas e litoral concentram os valores mais altos de transações. A dimensão das famílias  $H_4$ , concluindo-se assim a necessidade das famílias em adquirir casas maiores à medida que a família cresce. Por último o tipo de posse da propriedade  $H_{10}$  que afeta negativamente o valor, demonstrando assim que é pouco provável as famílias adquirirem uma segunda habitação, quando já são proprietários.

Posteriormente procedeu-se a uma análise de *clusters*, com o objectivo de verificar a existência de semelhanças entre municípios, no que respeita aos fatores que influenciam os valores da habitação em Portugal e averiguar se essas diferenças/semelhanças se manifestam efetivamente no VMPT. Da análise de *clusters* efetuada, concluiu-se que quando se consideram 2 *clusters*, o *cluster* 1 é composto pelos municípios com densidade populacional abaixo da média, enquanto que o *cluster* 2 é composto pelos municípios com maior densidade populacional. O *cluster* 2 é constituído por todos os municípios da área metropolitana do Porto e Lisboa e a maioria das capitais de distrito. O *cluster* 1 é composto por 261 municípios e parece corresponder aos municípios mais pobres e menos populosos, enquanto que o *cluster* 2 é composto por 47 municípios, demonstrando uma grande disparidade entre eles, parecendo corresponder aos municípios mais ricos e mais populosos. Podem ainda considerar-se 3 ou 4 *clusters*, destacando-se sempre um grupo de municípios com o VMHT muito mais alto do que os outros.

Este trabalho tem assim um contributo teórico, na medida em que se foca num estudo inter-municipal, o que não é comum na literatura, podendo apoiar a decisão municipal a adotar práticas de apoio à habitação. Os resultados obtidos permitem ainda uma aplicação prática, já que permite ao comprador conhecer as características dos municípios com valores de imóveis superiores.

Como limitações ao trabalho pode indicar-se o facto de os dados se restringirem ao ano de 2017, demonstrando alguma fragilidade quando se pretende extrapolar para outros anos. Ainda o facto dos Censos só se realizarem de 10 em 10 anos, pelo que se trabalhou com os dados disponíveis (2011).

Como sugestões futuras, seria interessante o estudo longitudinal, no sentido de confirmar os resultados ao longo de vários anos, assim como a inclusão de variáveis económicas como inflação, taxas de juro, PIB, entre outras. Para tal seria necessário recorrer a outra fonte

de dados já que o PORDATA não os disponibiliza, ou sugerir a sua inclusão nesta base de dados.



## Bibliografia

- Adair, Alastair S, Jim N Berry e W Stanley McGreal (1996). «Hedonic modelling, housing submarkets and residential valuation». Em: *Journal of property Research* 13.1, pp. 67–83.
- Alkana, Leyla et al. (2015). «Housing market differentiation: the cases of Yenimahalle and Çankaya in Ankara». Em: *International Journal of Strategic Property Management* 19.1, pp. 13–26.
- Alves e Pinheiro (2017). *O Setor Imobiliário em Portugal*. [https://www.bancobpi.pt/nocachecontent/conn/UCM/uuid/dDocName:PR\\_WCS01\\_UCM01061085](https://www.bancobpi.pt/nocachecontent/conn/UCM/uuid/dDocName:PR_WCS01_UCM01061085). Accessed: 2020-01-15.
- Archer, Wayne R, David C Ling e Brent C Smith (2010). «Ownership duration in the residential housing market: the influence of structure, tenure, household and neighborhood factors». Em: *The Journal of Real Estate Finance and Economics* 40.1, pp. 41–61.
- Banco de Portugal (2019). *Relatório de estabilidade financeira*. Accessed: 2020-01-21. url: [https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/ref\\_06\\_2018\\_pt.pdf](https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/ref_06_2018_pt.pdf).
- (2020). *Relatório de estabilidade financeira*. Accessed: 2020-12-12. url: [https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/ref\\_06\\_2020\\_pt.pdf](https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/ref_06_2020_pt.pdf).
- Bayet, Alain et al. (1991). «Les choix de consommation des ménages: une question de revenu avant tout». Em:
- Borio, Claudio EV e Patrick McGuire (2004). «Twin peaks in equity and housing prices?». Em: *BIS Quarterly Review*.
- Brasington, David M, Diane Hite e Andres Jauregui (2015). «House price impacts of racial, income, education, and age neighborhood segregation». Em: *Journal of Regional Science* 55.3, pp. 442–467.
- Calcagno, Riccardo, Elsa Fornero e MC Rossi (2009). «The effect of house prices on household consumption in Italy». Em: *The Journal of Real Estate Finance and Economics* 39.3, p. 284.
- Calmasur, Gurkan (2016). «Determining Factors Affecting Housing Prices in Turkey with Hedonic Pricing Model». Em: *International Conference on Business and Economics Studies, Washington DC, USA*, pp. 255–269.
- Chasco Yrigoyen, C e B Sánchez Reyes (2012). «Externalidades ambientales y precio de la vivienda en Madrid: Un análisis con regresión cuantílica espacial». Em: *Revista Galega de Economía* 21, pp. 1–21.
- Chevan, Albert (1982). «Age, housing choice, and neighborhood age structure». Em: *American Journal of Sociology* 87.5, pp. 1133–1149.
- Chong, Terence Tai Leung e Xiaoyang Li (2019). «The development of Hong Kong housing market: past, present and future». Em: *Economic and Political Studies*, pp. 1–20.
- Costa e Silva, E., A. Correia e A. Borges (2021). «Unveiling the Dynamics of the European Entrepreneurial Framework Conditions over the Last Two Decades: A Cluster Analysis». Em: *Axioms – aceite para publicação*.
- Duffie, Darrell (2019). «Prone to fail: The pre-crisis financial system». Em: *Journal of Economic Perspectives* 33.1, pp. 81–106.

- Eichholtz, Piet e Thies Lindenthal (2014). «Demographics, human capital, and the demand for housing». Em: *Journal of housing economics* 26, pp. 19–32.
- Ermisch, John (1996). «The demand for housing in Britain and population ageing: microeconomic evidence». Em: *Economica*, pp. 383–404.
- Fitch Osuna, Jesús Manuel, Karina Soto Canales e Ricardo Garza Mendiola (2013). «Valuación de la calidad urbano-ambiental. Una modelación hedónica: San Nicolás de los Garza, México». Em: *Estudios demográficos y urbanos* 28.2, pp. 383–428.
- Fuller, Gregory W, Alison Johnston e Aidan Regan (2020). «Housing prices and wealth inequality in Western Europe». Em: *West European Politics* 43.2, pp. 297–320.
- Gau, George W e Ko Wang (1994). «The tax-induced holding periods of real estate investors: theory and empirical evidence». Em: *The Journal of Real Estate Finance and Economics* 8.1, pp. 71–85.
- Guerreiro, António Eugénio Rocheta (2012). «Determinantes no preço de um imóvel: uma análise hedónica no contexto algarvio». Tese de mestrado.
- Gupta, Rangan e Alain Kabundi (2010). «The effect of monetary policy on house price inflation». Em: *Journal of Economic Studies*.
- Gyourko, Joseph e Peter Linneman (1996). «Analysis of the changing influences on traditional households' ownership patterns». Em: *Journal of Urban Economics* 39.3, pp. 318–341.
- Hair, Joseph et al. (2010). *Multivariate Data Analysis*. 7. New Jersey Pearson.
- Hoffman, Edward (1988). *The right to be human: A biography of Abraham Maslow*. JP Tarcher Los Angeles, CA.
- Huang, Youqin e William A. V. Clark (2002). «Housing Tenure Choice in Transitional Urban China: A Multilevel Analysis». Em: *Urban Studies* 39.1, pp. 7–32.
- Hui, Eddie CM et al. (2007). «Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix». Em: *Building and environment* 42.6, pp. 2333–2343.
- Idealista (2019). *Nos últimos 2 anos notamos uma tendência muito maior para os reembolsos antecipados*. Accessed: 2020-01-21. url: <https://www.idealista.pt/news/financas/credito-a-habitacao/2019/09/09/40783-uci>.
- Johnes, Geraint e Thomas Hyclak (1999). «House prices and regional labor markets». Em: *The Annals of Regional Science* 33.1, pp. 33–49.
- Katz, Lawrence e Kenneth T Rosen (1987). «The interjurisdictional effects of growth controls on housing prices». Em: *The Journal of Law and Economics* 30.1, pp. 149–160.
- Kaya, Asli e Murat Atan (2014). «Determination of the factors that affect house prices in Turkey by using hedonic pricing model». Em: *Journal of Business Economics and Finance* 3.3, pp. 313–327.
- Kotler, Philip e Gary Armstrong (2000). *Introdução ao marketing*. Livros Tecnicos e Científicos.
- Liu, Yan e Zhichao Li (2018). «Determinants of Housing Purchase Decision: An Empirical Study of the High Education Cohort in Urban China». Em: *Journal of Asian Architecture and Building Engineering* 17.2, pp. 299–305.
- Marôco, João (2010). «Análise Estatística com o SPSS statistics». Em: *ReportNumber, Lda*.
- McGreal, Stanley e Paloma Taltavull de La Paz (2013). «Implicit house prices: Variation over time and space in Spain». Em: *Urban Studies* 50.10, pp. 2024–2043.
- Moreira, Marta Susana Gomes (2000). «A dinâmica pública local e o valor da habitação: uma aplicação à área metropolitana do Porto». Em: *repositorio-aberto.up.pt*.

- Peng, Wensheng (2003). «Real estate indicators in Hong Kong». Em: *Conference on Real Estate Indicators and Financial Stability*. Vol. 27, p. 28.
- Rappaport, Jordan (2008). «Consumption amenities and city population density». Em: *Regional Science and Urban Economics* 38.6, pp. 533–552.
- Reis, Elizabeth (2001). «Estatística multivariada aplicada». Em: *Edições Sílabo*.
- República Portuguesa (1976). *Constituição da República portuguesa, VII Revisão*. Lisboa.
- Robson, Martin T (2003). «Housing markets and regional unemployment flows in Great Britain». Em: *The Manchester School* 71.2, pp. 132–155.
- Song, Yan e Gerrit-Jan Knaap (2003). «New urbanism and housing values: a disaggregate assessment». Em: *Journal of Urban Economics* 54.2, pp. 218–238.
- Tsai, I Chun e Peng (2011). «Bubbles in the Taiwan housing market: The determinants and effects». Em: *Habitat International* 35.2, pp. 379–390.
- Tsatsaronis, Kostas e Haibin Zhu (2004). «What drives housing price dynamics: cross-country evidence». Em: *BIS Quarterly Review, March*.
- Varão, Manuel Maria Góis Ferreira Proença (2019). «The determinants of housing prices in Portugal between 2000 and 2015». Tese de doutoramento.
- Winston, Nessa e Montserrat Pareja Eastaway (2008). «Sustainable housing in the urban context: international sustainable development indicator sets and housing». Em: *Social Indicators Research* 87.2, pp. 211–221.
- Xiao, Yue, Eddie C.M. Hui e Haizhen Wen (2020). «The housing market impacts of human activities in public spaces: The case of the square dancing». Em: *Urban Forestry Urban Greening* 54, p. 126769.
- Xu, Ting (2017). «The relationship between interest rates, Income, GDP growth and house prices». Em: *Research in Economics and Management* 2.1, pp. 30–37.
- Yanan, Song (2017). «The Impact of Demographic Structure on Housing Price: Evidence from Macau». Em: *Proceedings of the 2017 International Conference on Economics, Finance and Statistics (ICEFS 2017)*. Ed. por Atlantis Press, pp. 268–272.
- Zambrano-Monserrate, Manuel A (2016). «Formación de los precios de alquiler de viviendas en Machala (Ecuador): análisis mediante el método de precios hedónicos». Em: *Cuadernos de Economía* 39.109, pp. 12–22.
- Zhou, Qian et al. (fev. de 2020). «Do housing prices promote total factor productivity? Evidence from spatial panel data models in explaining the mediating role of population density». Em: *Land Use Policy* 91, p. 104410. doi: 10.1016/j.landusepol.2019.104410.



## **Apêndice A**

# **Distribuição dos Municípios por NUTS III**

<b>NUTSII</b>	<b>Frequência</b>	<b>Percentagem</b>
Alentejo Centro	14	4,5
Alentejo Litoral	5	1,6
Algarve	16	5,2
Alto Alentejo	15	4,9
Alto Minho	10	3,2
Alto Tâmega	6	1,9
Área Metropolitana de Lisboa	18	5,8
Área Metropolitana do Porto	17	5,5
Ave	8	2,6
Baixo Alentejo	13	4,2
Beira Baixa	6	1,9
Beiras e Serra da Estrela	15	4,9
Cávado	6	1,9
Douro	19	6,2
Ilha da Madeira	10	3,2
Ilha das Flores	2	0,6
Ilha de Porto Santo	1	0,3
Ilha de Santa Maria	1	0,3
Ilha de São Jorge	2	0,6
Ilha de São Miguel	6	1,9
Ilha do Corvo	1	0,3
Ilha do Faial	1	0,3
Ilha do Pico	3	1,0
Ilha Graciosa	1	0,3
Ilha Terceira	2	0,6
Lezíria do Tejo	11	3,6
Médio Tejo	13	4,2
Oeste	12	3,9
Região de Aveiro	11	3,6
Região de Coimbra	19	6,2
Região de Leiria	10	3,2
Tâmega e Sousa	11	3,6
Terras de Trás-os-Montes	9	2,9
Viseu Dão Lafões	14	4,5
<b>Total</b>	<b>308</b>	<b>100,0</b>

Tabela A.1: Distribuição dos Municípios por NUTS III

## **Apêndice B**

# **Dendrograma**

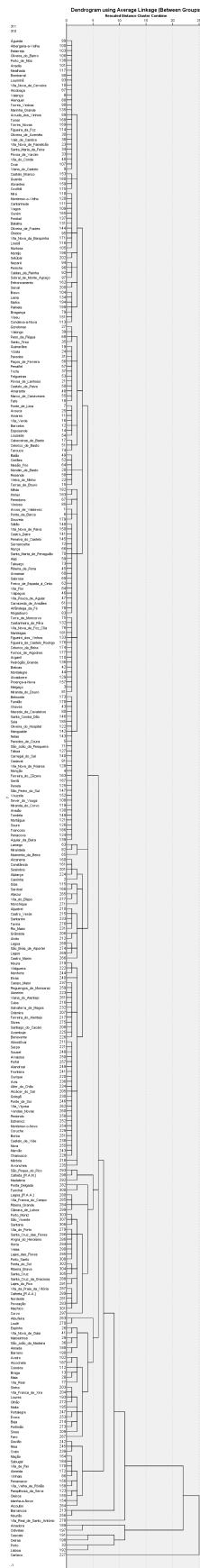


Figura B.1: Dendrograma

## **Apêndice C**

# **Comparação entre médias por Cluster do VMPT e das outras variáveis**

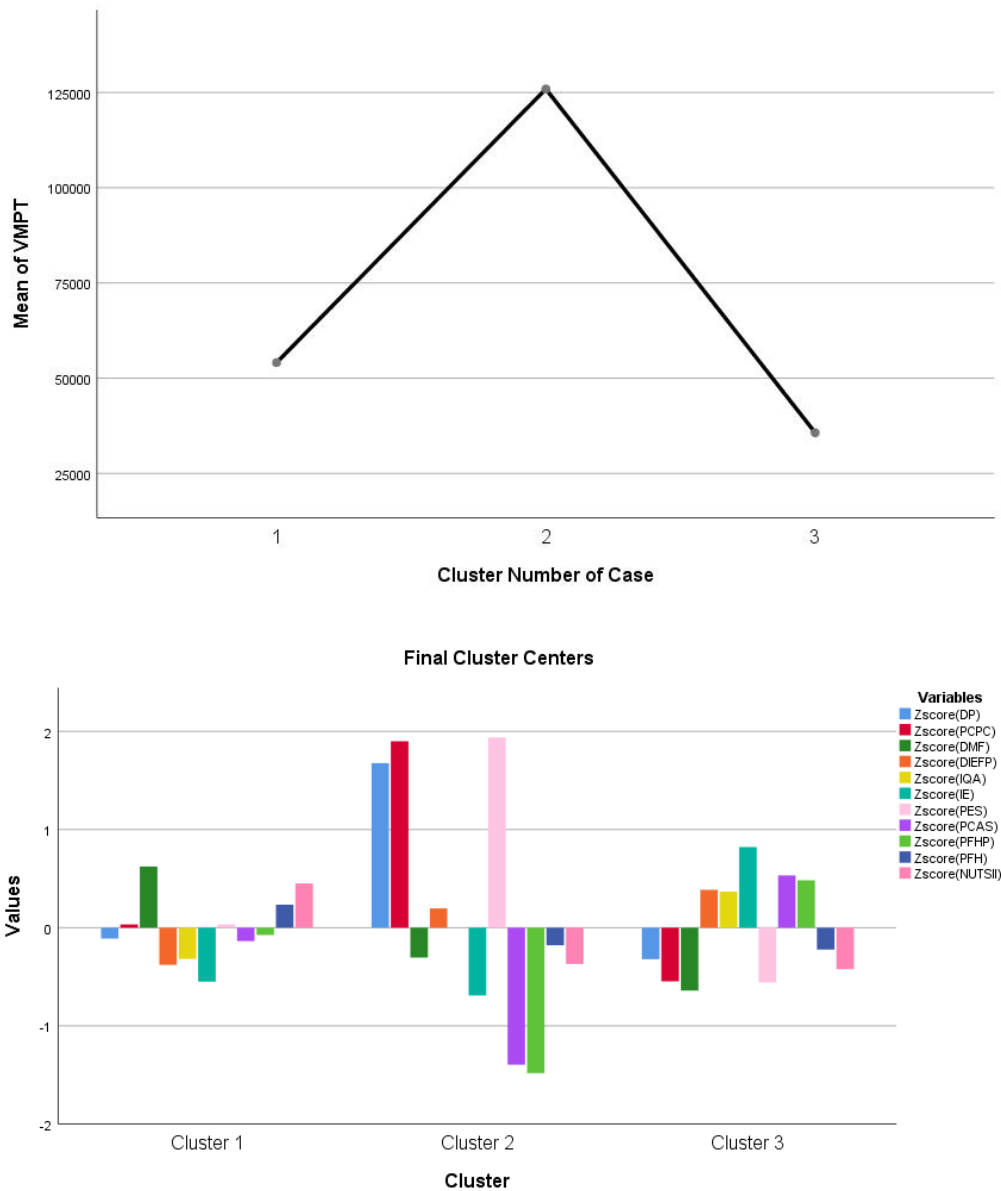


Figura C.1: Médias por cluster para  $k = 3$

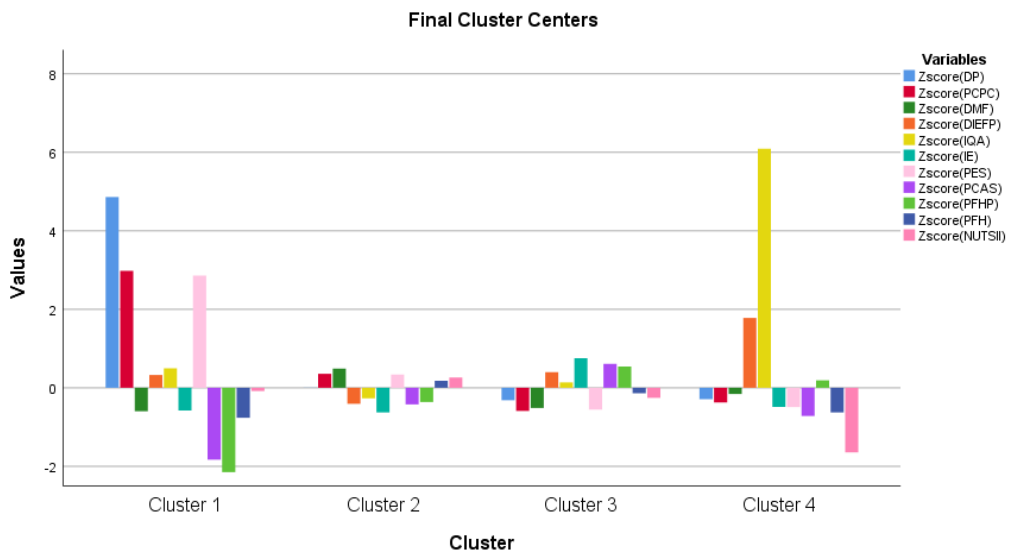
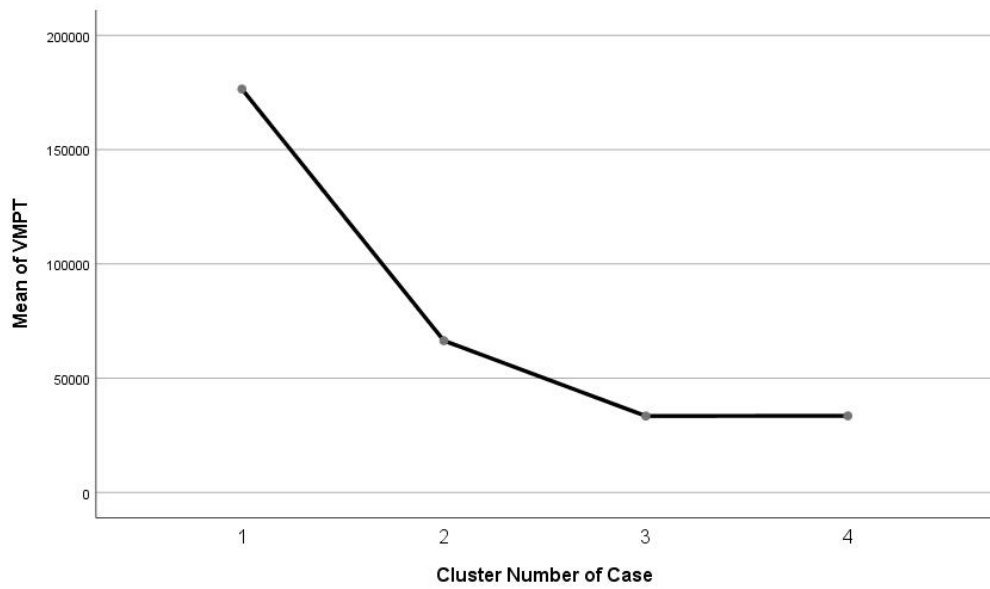


Figura C.2: Médias por cluster para  $k = 4$



## Apêndice D

# Testes ANOVA para 3 e 4 Clusters

	ANOVA					
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(DP)	55,304	2	,644	305	85,887	,000
Zscore(PCPC)	80,390	2	,479	305	167,686	,000
Zscore(DMF)	56,126	2	,639	305	87,900	,000
Zscore(DIEFP)	20,651	2	,871	305	23,705	,000
Zscore(IQA)	16,049	2	,901	305	17,806	,000
Zscore(IE)	73,145	2	,527	305	138,816	,000
Zscore(PES)	83,632	2	,458	305	182,542	,000
Zscore(PCAS)	52,498	2	,662	305	79,265	,000
Zscore(PFHP)	52,551	2	,662	305	79,386	,000
Zscore(PFH)	7,701	2	,956	305	8,055	,000
Zscore(NUTSII)	28,659	2	,819	305	35,009	,000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Figura D.1: Testes ANOVA para  $k = 3$  clusters

	ANOVA					
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(DP)	75,595	3	,264	304	286,497	,000
Zscore(PCPC)	49,352	3	,523	304	94,392	,000
Zscore(DMF)	25,917	3	,754	304	34,368	,000
Zscore(DIEFP)	19,553	3	,817	304	23,935	,000
Zscore(IQA)	42,457	3	,591	304	71,854	,000
Zscore(IE)	48,143	3	,535	304	90,026	,000
Zscore(PES)	45,036	3	,565	304	79,649	,000
Zscore(PCAS)	37,179	3	,643	304	57,823	,000
Zscore(PFHP)	34,384	3	,671	304	51,278	,000
Zscore(PFH)	4,753	3	,963	304	4,936	,002
Zscore(NUTSII)	9,314	3	,918	304	10,147	,000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Figura D.2: Testes ANOVA para  $k = 4$  clusters



## Apêndice E

### Elementos do Cluster 2, para $k = 2$

Os restantes concelhos pertencem ao *Cluster* 1.

Case Number	Município	Cluster	Distance
263	Albufeira	2	3,67
187	Alcochete	2	2,04
188	Almada	2	1,65
189	Amadora	2	7,48
102	Aveiro	2	2,34
190	Barreiro	2	2,10
214	Beja	2	2,48
13	Braga	2	2,42
92	Caldas da Rainha	2	2,50
191	Cascais	2	4,11
112	Coimbra	2	3,22
240	Elvas	2	3,74
162	Entroncamento	2	2,21
26	Espinho	2	3,00
253	Évora	2	2,40
267	Faro	2	2,92
300	Funchal	2	3,12
27	Gondomar	2	3,14
104	Ílhavo	2	2,18
268	Lagoa	2	3,41
269	Lagos	2	4,01
192	Lisboa	2	9,88
270	Loulé	2	3,32
193	Loures	2	1,56
194	Mafra	2	2,22
28	Maia	2	2,13
29	Matosinhos	2	2,54
195	Moita	2	3,02
196	Montijo	2	2,65
197	Odivelas	2	5,22
198	Oeiras	2	5,18
272	Olhão	2	2,85
199	Palmela	2	2,25
282	Ponta Delgada	2	4,20
247	Portalegre	2	2,89
273	Portimão	2	2,30
32	Porto	2	7,42
233	Santarém	2	2,56
36	São João da Madeira	2	2,77
200	Seixal	2	2,22
202	Setúbal	2	2,15
209	Sines	2	3,11
203	Sintra	2	1,93
39	Valongo	2	2,81
204	Vila Franca de Xira	2	2,25
41	Vila Nova de Gaia	2	2,63
77	Vila Real	2	3,13

Tabela E.1: Municípios no *Cluster* 2, para  $k = 2$



## Apêndice F

### Elementos do Cluster 2 e 3, para $k = 3$

Os restantes concelhos pertencem ao *Cluster* 1.

Case Number	Município	Cluster	Distance	Case Number	Município	Cluster	Distance
13	Braga	2	2,680	1	Arcos de Valdevez	3	2,381
26	Espinho	2	2,998	3	Melgaço	3	2,676
28	Maia	2	2,203	4	Monção	3	2,136
29	Matosinhos	2	2,192	5	Paredes de Coura	3	2,144
32	Porto	2	6,874	42	Boticas	3	1,993
36	São João da Madeira	2	2,551	44	Montalegre	3	2,139
41	Vila Nova de Gaia	2	2,693	45	Ribeira de Pena	3	1,944
102	Aveiro	2	2,499	46	Valpaços	3	2,042
112	Coimbra	2	3,178	47	Vila Pouca de Aguiar	3	1,734
162	Entroncamento	2	2,531	59	Alijó	3	1,964
187	Alcochete	2	2,330	60	Armamar	3	2,301
188	Almada	2	1,256	61	Carrazeda de Ansiães	3	1,938
189	Amadora	2	7,112	62	Freixo de Espada à Cinta	3	2,147
190	Barreiro	2	2,111	64	Mesão Frio	3	3,651
191	Cascais	2	3,821	66	Murça	3	2,093
192	Lisboa	2	9,314	67	Penedono	3	2,047
193	Loures	2	1,796	69	Sabrosa	3	1,746
196	Montijo	2	3,042	70	Santa Marta de Penaguião	3	2,552
197	Odivelas	2	4,970	72	Sernancelhe	3	2,125
198	Oeiras	2	4,683	73	Tabuaço	3	2,472
200	Seixal	2	2,588	75	Torre de Moncorvo	3	2,260
202	Setúbal	2	2,510	76	Vila Nova de Foz Côa	3	1,870
203	Sintra	2	2,269	78	Alfândega da Fé	3	2,369
204	Vila Franca de Xira	2	2,672	80	Macedo de Cavaleiros	3	1,791
209	Sines	2	3,364	81	Miranda do Douro	3	2,143
214	Beja	2	2,840	83	Mogadouro	3	2,203
233	Santarém	2	2,994	84	Vila Flor	3	1,802
247	Portalegre	2	3,157	85	Vimioso	3	2,777
253	Évora	2	2,567	86	Vinhais	3	3,281
263	Albufeira	2	3,823	110	Arganil	3	1,002
267	Faro	2	2,873	115	Góis	3	2,850
270	Loulé	2	3,602	121	Mortágua	3	1,697
273	Portimão	2	2,614	122	Oliveira do Hospital	3	1,744
300	Funchal	2	3,229	123	Pampilhosa da Serra	3	3,657
				124	Penacova	3	2,268
				125	Penela	3	1,798
				126	Soure	3	1,859
				129	Alvaiázere	3	1,974
				130	Ansião	3	1,838
				132	Castanheira de Pêra	3	2,069
				133	Figueiró dos Vinhos	3	1,187
				136	Pedrogão Grande	3	1,600
				139	Aguiar da Beira	3	2,017
				140	Carregal do Sal	3	1,793
				141	Castro Daire	3	1,992
				145	Penalva do Castelo	3	2,285
				146	Santa Comba Dão	3	1,611
				147	São Pedro do Sul	3	1,447
				148	Sátão	3	2,684
				149	Tondela	3	1,891
				150	Vila Nova de Paiva	3	1,977
				152	Vouzela	3	1,933
				154	Idanha-a-Nova	3	2,722
				155	Oleiros	3	4,126
				156	Penamacor	3	3,810
				157	Proença-a-Nova	3	1,731
				158	Vila Velha de Ródão	3	4,584
				159	Abrantes	3	2,243
				161	Constância	3	2,896
				163	Ferreira do Zêzere	3	1,655
				164	Mação	3	2,238
				166	Sardoal	3	2,737
				167	Sertã	3	1,214
				170	Vila de Rei	3	2,659
				172	Almeida	3	3,172
				173	Belmonte	3	1,547

(continua)

Tabela F.1: Municípios no Clusters 2 e 3, para  $k = 3$

Case Number	Município	Cluster	Distance	Case Number	Município	Cluster	Distance
					<b>(continuação)</b>		
				174	Celorico da Beira	3	1,483
				176	Figueira de Castelo Rodrigo	3	0,908
				177	Fornos de Algodres	3	1,374
				178	Fundão	3	1,563
				179	Gouveia	3	2,357
				181	Manteigas	3	1,389
				182	Mêda	3	2,328
				183	Pinhel	3	2,294
				184	Sabugal	3	2,548
				185	Seia	3	1,327
				186	Trancoso	3	1,847
				205	Alcácer do Sal	3	2,488
				207	Odemira	3	2,935
				208	Santiago do Cacém	3	2,888
				210	Aljustrel	3	2,313
				211	Almodôvar	3	1,589
				212	Alvito	3	3,115
				213	Barrancos	3	5,601
				216	Cuba	3	2,170
				217	Ferreira do Alentejo	3	2,259
				218	Mértola	3	2,181
				219	Moura	3	3,402
				220	Ourique	3	2,347
				221	Serpa	3	1,965
				222	Vidigueira	3	2,819
				224	Alpiarça	3	2,330
				227	Cartaxo	3	10,073
				228	Chamusca	3	2,378
				229	Coruche	3	1,870
				230	Golegã	3	2,457
				232	Salvaterra de Magos	3	2,374
				234	Alter do Chão	3	2,321
				235	Arronches	3	2,715
				236	Avis	3	2,263
				238	Castelo de Vide	3	2,151
				239	Crato	3	1,940
				240	Elvas	3	3,995
				241	Fronteira	3	1,646
				242	Gavião	3	3,100
				243	Marvão	3	2,588
				244	Monforte	3	3,054
				245	Nisa	3	2,296
				246	Ponte de Sor	3	1,985
				248	Sousel	3	1,770
				249	Alandroal	3	1,842
				250	Arraiolos	3	1,798
				251	Borba	3	2,149
				252	Estremoz	3	2,732
				254	Montemor-o-Novo	3	3,218
				255	Mora	3	2,086
				256	Mourão	3	4,733
				257	Portel	3	1,612
				258	Redondo	3	2,731
				259	Reguengos de Monsaraz	3	2,284
				261	Viana do Alentejo	3	2,526
				262	Vila Viçosa	3	2,534
				264	Alcoutim	3	4,215
				265	Aljezur	3	2,690
				266	Castro Marim	3	3,277
				271	Monchique	3	2,960
				277	Vila do Bispo	3	3,681

Tabela F.2: Municípios no Clusters 2 e 3, para  $k = 3$



## Apêndice G

# Elementos do Cluster 1, 3 e 4, para $k = 4$

Os restantes concelhos pertencem ao *Cluster 2*.

Case Number	Município	Cluster	Distance	Case Number	Município	Cluster	Distance	Case Number	Município	Cluster	Distance
29	Matosinhos	1	2,736	1	Arcos de Valdevez	3	2,143	213	Barrancos	4	2,612
32	Porto	1	3,816	3	Melgaço	3	2,544	227	Cartaxo	4	4,885
36	São João da Madeira	1	3,338	4	Monção	3	1,893	256	Mourão	4	2,475
188	Almada	1	2,924	5	Paredes de Coura	3	1,810				
189	Amadora	1	4,619	15	Terras de Bouro	3	2,776				
191	Cascais	1	4,024	20	Mondim de Basto	3	2,434				
192	Lisboa	1	6,261	22	Vieira do Minho	3	2,864				
197	Odivelas	1	3,842	42	Boticas	3	1,854				
198	Oeiras	1	2,889	43	Chaves	3	1,887				
				44	Montalegre	3	2,038				
				45	Ribeira de Pena	3	1,632				
				46	Valpaços	3	1,846				
				47	Vila Pouca de Aguiar	3	1,424				
				49	Baião	3	2,985				
				52	Cinfães	3	3,084				
				58	Resende	3	2,703				
				59	Alijo	3	1,718				
				60	Armamar	3	1,987				
				61	Carraceda de Ansiães	3	1,768				
				62	Freixo de Espada à Cinta	3	2,041				
				64	Mesão Frio	3	3,446				
				65	Moimenta da Beira	3	4,081				
				66	Murça	3	1,829				
				67	Penedono	3	1,923				
				69	Sabrosa	3	1,495				
				70	Santa Marta de Penaguião	3	2,242				
				71	São João da Pesqueira	3	1,816				
				72	Sernancelhe	3	1,783				
				73	Tabuaco	3	2,257				
				74	Tarouca	3	2,949				
				75	Torre de Moncorvo	3	2,269				
				76	Vila Nova de Foz Côa	3	1,779				
				78	Alfândega da Fé	3	2,212				
				80	Macedo de Cavaleiros	3	1,513				
				81	Miranda do Douro	3	2,097				
				82	Mirandela	3	2,982				
				83	Mogadouro	3	2,110				
				84	Vila Flor	3	1,717				
				85	Vimioso	3	2,695				
				86	Vinhais	3	3,289				
				91	Cadaval	3	1,473				
				108	Sever do Vouga	3	1,697				
				110	Arganil	3	0,752				
				115	Góis	3	3,074				
				118	Mira	3	2,251				
				119	Miranda do Corvo	3	1,750				
				120	Montemor-o-Velho	3	2,439				
				121	Mortágua	3	1,508				
				122	Oliveira do Hospital	3	1,471				
				123	Pampilhosa da Serra	3	3,819				
				124	Penacova	3	1,962				
				125	Penela	3	1,759				
				126	Soure	3	1,583				
				127	Tábua	3	1,624				
				129	Alvaiázere	3	1,814				
				130	Ansião	3	1,525				
				132	Castanheira de Pêra	3	2,096				
				133	Figueiro dos Vinhos	3	1,071				
				136	Pedregão Grande	3	1,486				
				139	Aguiar da Beira	3	1,809				
				140	Carregal do Sal	3	1,484				
				141	Castro Daire	3	1,684				

Tabela G.1: Municípios no *Cluster 1, 3 e 4*, para  $k = 4$

Case Number	Município	Cluster	Distance	Case Number	Município	Cluster	Distance	Case Number	Município	Cluster	Distance
					<b>(Continuação)</b>						
				142	Mangualde	3	1,755				
				143	Nelas	3	1,719				
				145	Penalva do Castelo	3	1,973				
				146	Santa Comba Dão	3	1,310				
				147	São Pedro do Sul	3	1,077				
				148	Sátão	3	2,414				
				149	Tondela	3	1,642				
				150	Vila Nova de Paiva	3	1,712				
				152	Vouzela	3	1,618				
				154	Idanha-a-Nova	3	2,773				
				155	Oleiros	3	4,169				
				156	Penamacor	3	3,904				
				157	Proença-a-Nova	3	1,665				
				158	Vila Velha de Ródão	3	4,676				
				159	Abrantes	3	2,087				
				163	Ferreira do Zêzere	3	1,464				
				164	Mação	3	2,276				
				166	Sardoal	3	2,925				
				167	Sertã	3	1,149				
				170	Vila de Rei	3	2,724				
				172	Almeida	3	3,332				
				173	Belmonte	3	1,493				
				174	Celorico da Beira	3	1,226				
				176	Figueira de Castelo Rodrigo	3	0,824				
				177	Fornos de Algodres	3	1,138				
				178	Fundão	3	1,355				
				179	Gouveia	3	2,225				
				181	Manteigas	3	1,314				
				182	Méda	3	2,213				
				183	Pinhel	3	2,130				
				184	Sabugal	3	2,625				
				185	Seia	3	1,102				
				186	Trancoso	3	1,635				
				205	Alcácer do Sal	3	2,661				
				207	Odemira	3	3,108				
				210	Aljustrel	3	2,450				
				211	Almodôvar	3	1,793				
				212	Alvito	3	3,198				
				216	Cuba	3	2,180				
				217	Ferreira do Alentejo	3	2,432				
				218	Mértola	3	2,408				
				219	Moura	3	3,427				
				220	Ourique	3	2,652				
				221	Serpa	3	2,132				
				222	Vidigueira	3	2,905				
				224	Alpiarça	3	2,576				
				228	Chamusca	3	2,542				
				229	Coruche	3	2,078				
				232	Salvaterra de Magos	3	2,314				
				234	Alter do Chão	3	2,654				
				235	Arronches	3	2,890				
				236	Avis	3	2,531				
				238	Castelo de Vide	3	2,430				
				239	Crato	3	2,208				
				241	Fronteira	3	1,823				
				242	Gavião	3	3,443				
				243	Marvão	3	2,836				
				244	Monforte	3	3,192				
				245	Nisa	3	2,620				
				246	Ponte de Sor	3	2,195				
				248	Sousel	3	2,009				
				249	Alandroal	3	1,905				
				250	Arraiolos	3	1,970				
				251	Borba	3	2,191				
				252	Estremoz	3	2,833				
				254	Montemor-o-Novo	3	3,318				
				255	Mora	3	2,328				
				257	Portel	3	1,817				
				258	Redondo	3	2,840				
				259	Reguengos de Monsaraz	3	2,315				
				261	Viana do Alentejo	3	2,534				
				262	Vila Viçosa	3	2,660				
				264	Alcoutim	3	4,338				
				265	Aljezur	3	2,952				
				266	Castro Marim	3	3,369				
				271	Monchique	3	3,318				
				277	Vila do Bispo	3	3,953				

Tabela G.2: Municípios no Cluster 1, 3 e 4, para  $k = 4$